7-20-2020

The Economics of Babysitting a Robot

Aleksandr Alekseev

The Economics of Babysitting a Robot

Comments

ESI Working Paper 20-29

The Economics of Babysitting a Robot *

Aleksandr Alekseev†

July 20, 2020

Abstract

I study the welfare effect of automation on workers in a setting where technology is complementary but imperfect. Using a modified task-based framework, I argue that imperfect complementary automation can impose non-pecuniary costs on workers via a behavioral channel. The theoretical model suggests that a critical factor determining the welfare effect of imperfect complementary automation is the automatability of the production process. I confirm the model's predictions in an experiment that elicits subjects' revealed preference for automation. Increasing automatability leads to a significant increase in the demand for automation. I explore additional drivers of the demand for automation using machine learning analysis and textual analysis of choice reasons. The analysis reveals that task enjoyment, performance, and cognitive flexibility are the most important predictors of subjects' choices. There is significant heterogeneity in how subjects evaluate imperfect complementary automation. I discuss the implications of my results for workers' welfare, technology adoption, and inequality.

Keywords: automation, worker welfare, imperfect technology, task-switching, personnel economics, experiment

JEL codes: C91, D63, D91, M52, J24, O33

*I thank James Bland, Klajdi Bregu, Gabriele Camera, Brice Corgnet, John Duffy, Joaquin Gómez-Miñambres, Glenn Harrison, Erik Kimbrough, Philipp Limberg, Nate Neligh, David Neumark, David Porter, David Rojo-Arjona, and Nat Wilcox for their helpful suggestions and comments on the earlier drafts of this paper. I thank seminar participants at UC Irvine and Chapman University (Economic Science Institute), as well as conference participants at the 2019 ESA North American Meeting and the Southwest Experimental and Behavioral Economics Workshop 2019 for their valuable feedback. Nannan Peng provided outstanding research assistance. Financial support from the Economic Science Institute at Chapman University is gratefully acknowledged.

†Economic Science Institute, Chapman University, One University Drive, Orange, CA, 92866, e-mail: alekseev@chapman.edu, ORCID: 0000-0001-6542-1920.

1 Introduction

Automation is one of the most important trends in the labor market that will have a long-lasting impact on firms, workers, and public policies.[1] This trend has initiated a growing literature in economics that studies the market-level effects of automation. The existing literature seeks to provide both a theoretical account of the potential implications of automation for growth and labor share (Frey and Osborne, 2017; Brynjolfsson, Rock, and Syverson, 2017; Acemoglu and Restrepo, 2018b, 2019; Agrawal, Gans, and Goldfarb, 2019), as well as estimates of the contemporary effects of automation on the labor markets in the US and Europe (Goos, Manning, and Salomons, 2014; Autor, 2015; Lordan and Neumark, 2018; Acemoglu and Restrepo, 2020). What the existing analysis tends to be missing, however, is the account of the micro-level, or behavioral, effects of automation.[2] How would it feel to work alongside a robot? How would it change the work environment? How would it change the incentives that workers face? These are equally important questions to answer given that an ever-increasing number of workers have to deal with automation in their workplaces. Answering these questions is necessary to get a complete picture of the potential effects of automation on workers' welfare and inequality.

I study the behavioral effects of automation in the workplace in a setting where automation technology is complementary to workers' skills. I argue that automation is not always beneficial for workers in this case, despite the common belief. My argument proceeds in two ways. First, I develop a theoretical model and use it to identify the conditions under which complementary automation benefits or hurts workers. Second, I conduct an experiment to empirically analyze the behavioral effects of automation and test the model's predictions. To the best of my knowledge, the present study provides the first causal evidence on the behavioral effects of complementary automation. I show that in some instances, complementary automation will reduce workers' welfare. Complementary automation can also make welfare inequality among different types of workers more pronounced. Having identified these issues, I propose several mechanisms to compensate workers for the adverse effects of automation.

My focus on complementary automation is motivated by the emerging consensus that whether

[1] Automation is, of course, not a new phenomenon. It has been occurring in waves throughout history. See Mokyr, Vickers, and Ziebarth (2015) for an overview of automation in a historical context.

[2] While the studies of the behavioral effects of automation in the workplace begin to emerge (Corgnet, Hernán-González, and Mateo, 2018; Granulo, Fuchs, and Puntoni, 2019), the evidence is still relatively scarce.

automation will benefit or hurt workers depends on whether automation complements or substitutes their skills (Autor, 2015; Agrawal, Gans, and Goldfarb, 2019). Automation that substitutes workers' skills will ultimately displace workers and reduce their welfare. Automation that complements workers' skills, on the other hand, should make workers more productive, which in turn should lead to higher wages and overall welfare. I ask if it is always true that complementary automation increases workers' welfare. I argue that the answer to this question is not always affirmative. The key reason why complementary automation might not always be beneficial for workers is that new technologies are often imperfect (Acemoglu and Restrepo, 2019).

A recent example of such complementary but imperfect automation comes from Walmart. In 2019, Walmart began expanding the use of cleaning robots in their stores to speed up cleaning. The company executives claimed that the addition of such robots would free employees from routine work and allow them to focus on more meaningful tasks. In reality, however, employees were infuriated about this new work environment. The employees were complaining that they had to manage a more tedious task under this new environment—babysitting robots by correcting their errors. For example, robots would often be stuck somewhere and send text messages to employees with requests of help.[3]

My theoretical model formalizes the intuition behind the Walmart example. On the one hand, automation is complementary, which allows workers to perform new better tasks and benefits them.[4] On the other hand, automation is imperfect in the sense that workers occasionally have to go back to old tasks and babysit the robot. Babysitting creates a task-switching environment and imposes non-pecuniary costs on workers via a behavioral channel.[5] The trade-off between these costs and benefits determines the net welfare effect of automation.[6] My modeling approach draws inspiration from the task-based framework (Zeira, 1998; Autor, Levy, and Murnane, 2003; Acemoglu and Autor,

[3]See, e.g., Kristin Houser (https://futurism.com/grouchy-employee-hates-walmart-robots), Drew Harwell (https://www.washingtonpost.com/technology/2019/06/06/walmart-turns-robots-its-human-workers-who-feel-like-machines/).

[4]In the terminology of Acemoglu and Restrepo (2018b), automation technology combines the process of the displacement of labor and new task creation.

[5]The negative effects of multitasking and task-switching are well-documented in the economics (Holmstrom and Milgrom, 1991; Dewatripont, Jewitt, and Tirole, 1999; Coviello, Ichino, and Persico, 2014) and psychology (Spector and Biederman, 1976; Kiesel et al., 2010; Vandierendonck, Liefooghe, and Verbruggen, 2010) literatures. On the other hand, some studies in organizational psychology document positive effects of task variety for some workers (Zaniboni, Truxillo, and Fraccaroli, 2013). I revisit this latter possibility when I analyze subjects' stated reasons for their choices.

[6]The costs and benefits here are viewed from the subjective perspective of workers. These costs and benefits will likely be distinct from the costs and benefits of automation from a firm's perspective.

2011; Acemoglu and Restrepo, 2018a) by assuming that the production process consists of a series of tasks assigned to an agent. However, a significant difference is that I allow tasks to differ along two dimensions: type and difficulty. My modeling of the task-switching environment is different from the approach taken in the principal-agent literature (Holmstrom and Milgrom, 1991; Dewatripont, Jewitt, and Tirole, 1999) in the sense that tasks are assigned sequentially and switching between task types is costly. The model highlights that a critical factor determining the net welfare effect of automation is the *automatability* of the environment, or how well the production process is suited for automation. I show that if automatability exceeds a certain threshold, the net effect of automation will be positive, and *vice versa*. I derive the comparative statics for the demand for automation in terms of the parameters of the work environment and agents' characteristics, which allows me to test the model's predictions experimentally.

The experiment elicits subjects' revealed preference for automation since one cannot directly observe the welfare effect of automation. I use a choice-from-experience design (Niederle and Vesterlund, 2007; Hertwig and Erev, 2009) in which subjects first experience two types of technologies, manual and automatic, and then choose the technology they prefer. Under manual technology, subjects work on their own. Under the automatic technology, subjects work with the computer. The computer allows subjects to work on a new task and earn more money, but occasionally it requires babysitting. The experimental task combines the elements of the real-effort framework with the task-switching framework (Kiesel et al., 2010; Vandierendonck, Liefooghe, and Verbruggen, 2010). Subjects' choices are incentivized, such that the choice of technology has real consequences for the payoffs that they earn. The primary outcome of interest is the aggregate preference, or demand, for the automatic technology. My primary treatment variable is the automatability of the environment. The experimental results confirm that automatability has a positive and significant causal effect on the demand for automation.

I supplement the causal analysis of treatment effects with an exploratory machine learning analysis in which I ask which subjects' characteristics could further explain the observed variation in choices. I consider a comprehensive suite of potential predictors, including demographics, personality traits, risk and time preferences, and task performance. To avoid overfitting, I use the Lasso estimator, a standard tool in machine learning that is gaining popularity in economics (Varian, 2014; Mullainathan and Spiess, 2017). The results show that the strongest predictors of subjects'

choices are measures of task enjoyment, performance, and cognitive flexibility.

In addition to the quantitative analysis of subjects' choices, I conduct a qualitative analysis by eliciting the reasons behind choices. Subjects provide their answers as a free-form text, which is then classified. Many of the stated reasons can be mapped into the motivations suggested by the theoretical model. However, some of the responses yield unexpected results. For example, some subjects indicate that they enjoy a task-switching environment, which suggests that they experience a positive effect of task variety instead of an adverse effect of task switching.[7] These results, together with the overall variation in subjects' choices, suggest that the benefits and costs of automation are likely to be heterogeneous.

I draw two broad policy implications from my analysis. The first policy implication is that the welfare effect of complementary automation is more complicated than suggested by the current literature. I show that if automation is imperfect, it will generate non-pecuniary costs for workers, which can potentially outweigh the benefits. My theoretical results clarify the conditions under which complementary automation benefits or hurts workers. These conditions can help companies implement automation in the workplace and policymakers to regulate the adoption of automation technology. The second implication is that different types of workers might benefit differently from complementary automation. The adoption of the automation technology will likely create winners and losers even in the complementary case, which will lead to an even greater welfare inequality among workers. Identifying which types of workers benefit more or less from automation and testing the effectiveness of the compensation mechanisms proposed by the theoretical model are promising avenues for further research.

Related Literature

Corgnet, Hernán-González, and Mateo (2018) study the effect of automation on social incentives in the workplace using an incentivized lab experiment. They show that replacing a human team member with a robot results in a reduction in performance of the remaining human team members. The negative effect of the human-robot replacement is attributed to reduced social incentives.

[7] The absence of task variety and resulting drudgery is a major issue in many low-skill occupations, e.g., warehouse work. Amazon, among other companies, is attempting to combat this issue using the gamification of the work process. See, e.g., Greg Bensinger (https://www.washingtonpost.com/technology/2019/05/21/missionracer-how-amazon-turned-tedium-warehouse-work-into-game/).

Social pressure and envy are highlighted as the key social incentives that drive their results.

Granulo, Fuchs, and Puntoni (2019) use surveys to gauge preferences for the replacement of a human worker by either another worker or a robot. A respondent's perspective is manipulated by framing the question such that the worker who is being replaced is either the respondent or a third person. They find that the respondents' stated preferences depend on the perspective. Respondents would prefer a worker to be replaced by another worker, if the worker who is being replaced is a third person. However, respondents would prefer a worker to be replaced by a robot, if the worker who is being replaced is the respondent. These results are attributed to the differences in perceived self-threat and future work concerns associated with each replacement option and perspective.

Acemoglu and Restrepo (2018a,b, 2019, 2020) develop a task-based framework[8] to analyze the effects of automation on the labor market and use the framework to empirically evaluate these effects in the US context. Acemoglu and Restrepo (2018a,b) use a task-based framework to analyze the effects of new technologies, such as automation, on the labor market. The key idea of their framework is to represent a production process as a set of tasks that can be assigned either to labor or capital. This framework allows for new technologies, such as automation, that are different from the standard factor-augmenting technologies. The authors then analyze different potential effects of these new technologies on the labor market and identify the conditions under which the labor market outcomes will deteriorate or improve. Acemoglu and Restrepo (2020) empirically evaluate the effect of robots on local labor markets in the US. Local labor markets with great exposure to robots are shown to have worse outcomes in terms of employment and wages. The effect of robotization is separated from other competing effects, such as offshoring and greater exposure to foreign imports. Acemoglu and Restrepo (2019) analyze the evolution of labor demand in the US using the task-based framework. The framework allows them to decompose the overall changes in labor demand into a variety of effects. Changes to the task content of production are claimed to play an important role in shaping labor demand. Their findings suggest that in recent decades the process of automation has been accelerating while the process of new tasks creation has been decelerating.

Aghion, Jones, and Jones (2017) develop a model of economic growth in the presence of automation technologies. They use their growth model to study the effects of automation on industrial

[8]Their framework is based on Acemoglu and Autor (2011), Autor, Levy, and Murnane (2003), and Zeira (1998).

organization and wage inequality. A key feature of their model is the introduction of Baumol's "cost decease" argument, predicting the existence of bottlenecks in automation. This argument suggests that economic growth is likely to be constrained not by sectors with high productivity growth, but rather by sectors that are hard to improve. One of the implications of the model is that, even in the presence of automation, the share of labor in production will not be driven to zero.

Agrawal, Gans, and Goldfarb (2019) discuss the potential labor market effects of new technologies in a specific context of prediction and decision-making tasks. They focus on machine learning, and related automation of prediction tasks, and ask how these changes could affect labor market outcomes. The key to their discussion is the effect of the automation of prediction tasks on the relative returns of capital versus labor in decision-making tasks.

Frey and Osborne (2017) empirically analyze the likelihood of computerization for a large set of occupations. They find that occupations related to service, sales, office and administrative support have a high likelihood of being computerized. Occupations related to management, education, engineering and healthcare have a relatively low likelihood of being computerized. The potential effects of computerization on labor market outcomes in these occupations are discussed. A negative relationship between the estimated likelihood of computerization and wage for a given occupation is reported.

2 Theoretical Framework

In this section I develop a stylized model, which captures the trade-off associated with imperfect complementary automation and elucidates the factors that affect this trade-off. My modeling approach follows the task-based framework of Acemoglu and Restrepo (2018b) by assuming that the production process consists of a series of tasks assigned to an agent. I model automation as a technology that allows the agent to delegate some of the tasks to the robot (computer, machine). Using the robot enables the agent to work on new tasks, which represents the complementarity of automation. The imperfection of the automation technology is modeled as the inability of the robot to complete certain tasks: the robot passes those tasks back to the agent, which creates a multitasking (or task-switching) environment. The modeling of the multitasking environment,

however, is different from the approach taken in the principal-agent literature (Holmstrom and Milgrom, 1991; Dewatripont, Jewitt, and Tirole, 1999), since in the present case the tasks are assigned sequentially and switching between task types is costly.

2.1 Environment

Time is discrete and has two periods indexed by $t \in \{0, 1\}$. In each period an agent is assigned a *task*. The first dimension along which tasks vary is type, $\eta_t \in H = \{A, B\}$. Tasks of different types require different skills to complete them.[9] The second dimension along which tasks vary is difficulty, $\theta_t \in \Theta = \{l, h\}$. Tasks with higher difficulty require more effort to complete them. A task is then represented by $x_t \in X = H \times \Theta$. The agent exerts effort $e_t \in \mathbb{R}_+$ to complete a task. The history of tasks and exerted effort up to period t is denoted χ_t.

The outcome of a task, success or failure, is deterministic conditional on exerted effort and the characteristics of the task. The outcome is given by $z_t(x_t, e_t) \equiv z(x_t, e_t \mid \chi_t) = \mathbb{1}(e_t \geqslant \bar{e}_t(x_t))$. The interpretation of $\bar{e}_t(x_t) \equiv \bar{e}(x_t \mid \chi_t)$ is that it is an *effort requirement* for a task: the minimum amount of effort needed to complete a task. The effort requirement evolves over time via a *learning-by-doing* effect (Arrow, 1971).

Specifically, I assume that the effort requirement decreases after a repeated assignment of a task of the same type:

$$\bar{e}_1(x_1) = (\bar{e}_0(x_1) - \lambda \min\{e_0, \bar{e}_0(x_0)\} \mathbb{1}(\eta_1 = \eta_0))^+ , \tag{1}$$

where $\bar{e}_0(x) \geqslant 0$ denotes the initial effort requirement for task x.[10] The assumption that switching between task types results in a higher effort requirement is supported by the literature on a task-switching paradigm in psychology (Spector and Biederman, 1976; Kiesel et al., 2010; Vandierendonck, Liefooghe, and Verbruggen, 2010). Parameter $\lambda \in [0, 1]$ determines the strength of learning. I choose an additive form of learning in favor of a more traditional power form of learning (Wright, 1936) for the sake of greater tractability: the case of full learning with $\lambda = 1$ allows one to significantly simplify some of the expressions. I denote the initial effort requirements for a task of

[9] In contrast to the multitasking models in the principal-agent literature, task types do not have an alternative interpretation of different work attributes, such as quantity and quality, since tasks are assigned sequentially.

[10] Expression $\min\{e_0, \bar{e}_0(x_0)\}$ guarantees that the agent has no incentive to "store" effort across periods. The exact meaning of this assumption will become clear when I turn to the optimal effort choice. The non-negativity qualification, $(\cdot)^+ \equiv \max\{\cdot, 0\}$, guarantees that the effort requirement does not fall below zero.

type A as $a^l \equiv \bar{e}_0(Al), a^h \equiv \bar{e}_0(Ah)$, and the difference as $\Delta a \equiv a^h - a^l$, with $a^h > a^l$. I assume that the effort requirement for a task of type B does not depend on difficulty, and denote it as $b \equiv \bar{e}_0(Bl) = \bar{e}_0(Bh)$.

2.2 Production Process

The production process can be organized using one of the two available technologies, or *modes*, denoted as $\mu \in \{\mathcal{A}, \mathcal{M}\}$, where $\mathcal{A}$ is an Automatic mode and $\mathcal{M}$ is a Manual mode. Modes determine the *task content of production*, represented by the probability distribution of tasks $\mathbb{P}(x_t \mid \mu)$, as well as payoffs to the agent. The task content of production in the Manual mode is

$$\mathbb{P}(x_t \mid \mathcal{M}) = \begin{cases} p, & x_t = Al, \\ 1-p, & x_t = Ah, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

In the Manual mode the agent only works on tasks of type A. Parameter $p \in [0,1]$ determines the frequency of easy tasks of this type. I refer to p as the *automatability* of the production process. The task content of production is independent of history.

In the Automatic mode tasks Al are automated by a robot. The robot always successfully completes tasks Al, however, it cannot complete tasks Ah, which represents an imperfection of the automation technology. When task Ah arrives, it is always passed back to the agent. While the robot works on task Al, the agent can work on a task of type B. The task content of production in the Automatic mode is, therefore,

$$\mathbb{P}(x_t \mid \mathcal{A}) = \begin{cases} p, & x_t = Bl, \\ 1-p, & x_t = Ah, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

The automatability of the environment affects the efficiency of the robot in the sense that the robot can solve most of the tasks of type A in an environment with high p . However, automatability is conceptually different from automation efficiency since automatability also affects the agent by

changing the task content of production.

The agent's payoff depends on the outcome of a task and mode: $\pi(x_t \mid z_t, \mu) = z_t \pi(x_t \mid \mu)$. In the Manual mode the payoff for the successful completion of a task is given by

$$\pi(x_t \mid \mathcal{M}) = \begin{cases} \pi^A, & x_t \in \{Al, Ah\}, \\ 0, & \text{otherwise.} \end{cases}$$

For simplicity, I assume that the payoff depends only on a task type but not on difficulty.

In the Automatic mode the payoff for the successful completion of a task is given by

$$\pi(x_t \mid \mathcal{A}) = \begin{cases} \pi^A + \pi^B, & x_t = Bl, \\ \pi^A, & x_t = Ah, \\ 0, & \text{otherwise.} \end{cases}$$

I assume that $\pi^A, \pi^B \in \mathbb{R}_+$. The payoffs should be interpreted broadly as comprised of both extrinsic monetary rewards and intrinsic utility (Benabou and Tirole, 2003). The extrinsic part can represent a piece-rate incentive scheme, as well as a wage-plus-bonus incentive scheme in which wage is conditional on satisfying work requirements. The increased payoff for a task of type B can represent both a higher extrinsic reward due to greater productivity of the agent working with the robot, as well as higher intrinsic utility if a task of type B is more desirable and prestigious than a task of type A. For example, this would be the case if tasks of type A are routine and tasks of type B are creative.

2.3 Preferences

The agent has preferences over payoffs π and effort e represented by a utility function $u_t : \mathbb{R}_+ \times \mathbb{R}_+ \mapsto \mathbb{R}$. I make standard assumptions about the utility function: it is strictly increasing and concave in payoffs, and strictly decreasing and concave in effort. I further assume that the utility function takes a standard additively separable form: $u_t(\pi, e) = \pi - c_t(e)$, where $c_t : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is a cost of effort function.[11] I assume that c_t is strictly increasing and convex, and that $c_t(0) = 0$. These

[11] The additively separable form, while being analytically convenient, is also somewhat restrictive because it does not allow for complementarity or substitutability of effort and money in the utility function (Alekseev, 2020).

assumptions imply that c_t is superadditive: $c_t(e + \Delta e) \geqslant c_t(e) + c_t(\Delta e)$. Superadditivity has an intuitive interpretation. Suppose that $e = a^l$ and $\Delta e = \Delta a$. Then superadditivity implies that the agent finds it less demanding to complete an easy task followed by a hard task (assuming full learning) rather than to complete a hard task by itself.

The cost function depends on the history via *task-switching*. I assume that switching to a different task type in period 1 induces a switching cost $s \in \mathbb{R}_+$ that is added to the cost of effort:

$$c_1(e) = c_0(e) + s\,\mathbb{1}(\eta_1 \neq \eta_0).$$

The switching cost can be interpreted as the loss of the agent's utility due to psychological effects, such as irritation or an activation of a costly cognitive control (Kiesel et al., 2010; Vandierendonck, Liefooghe, and Verbruggen, 2010). This effect is conceptually different from the effect of task switching on effort requirement via the learning channel (1). The agent's utility of exerting effort e on a task x_t in period t is

$$U(e \mid x_t, \chi_t, \mu) = \pi(x_t \mid z_t, \mu) - c_t(e) = \mathbb{1}(e \geqslant \bar{e}_t(x_t))\pi(x_t \mid \mu) - c_t(e).$$

I assume that the participation constraint is satisfied: $\pi(x_t \mid \mu) > c_t(\bar{e}_t(x_t))$. There is no discounting between the periods. The agent aggregates utilities across periods and states according to the Expected Utility.

2.4 Optimal Choice

I now proceed to deriving the values of the Manual and Automatic modes to the agent, $V(\mu)$. Since these values are conditional on the agent's optimal behavior, I begin with the following proposition.[12]

Proposition 1. *The agent's optimal effort on task x in period t is the effort requirement of task x in that period.*

This result is driven by two facts. First, the participation constraint makes it sub-optimal for the agent to exert less effort than required because the agent can always get a strictly higher utility

[12] All proofs and derivations are in Appendix A.

by exerting exactly the effort required. Second, the learning process (1) caps the reduction of the effort requirement that can be achieved by exerting effort in period 0. This prevents the agent from "storing" effort by exerting higher-than-required effort in period 0 in order to lower the effort requirement in period 1 even further and get higher utility. This result implies that the agent's value in period t is

$$U^*(x_t, \chi_t, \mu) = \pi(x_t \mid \mu) - c_t(\bar{e}_t(x_t)).$$

Figure 1 presents a scheme of state-contingent values in each period under the Manual mode. Using this scheme, it is straightforward to compute the agent's value of the Manual mode. The value of the Manual mode under full learning, $\lambda = 1$, is[13]

$$V(\mathcal{M} \mid \lambda = 1) = 2\pi^A - \Big(pc_0(a^l) + (1-p)c_0(a^h) + p(1-p)c_0(\Delta a)\Big). \tag{4}$$

Figure 1: Scheme of Manual Mode

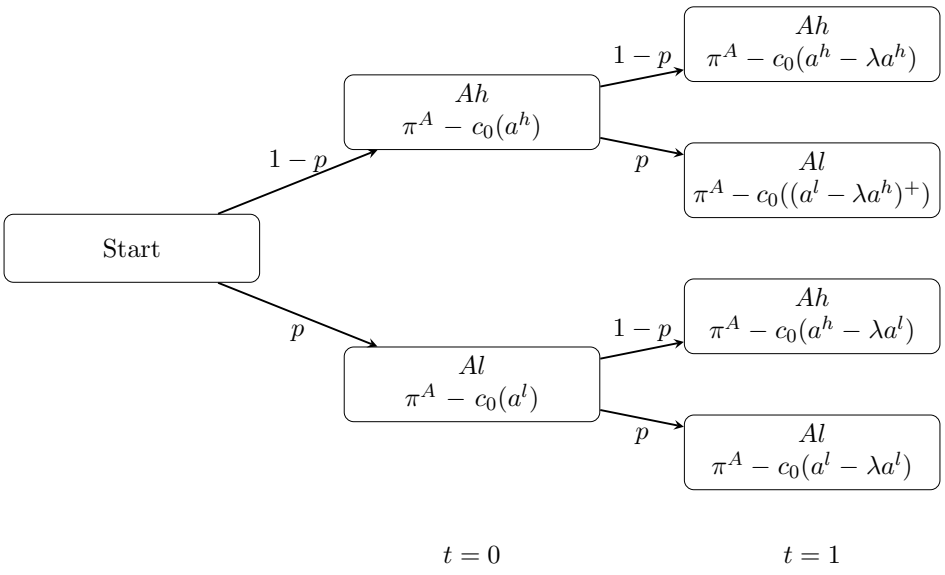


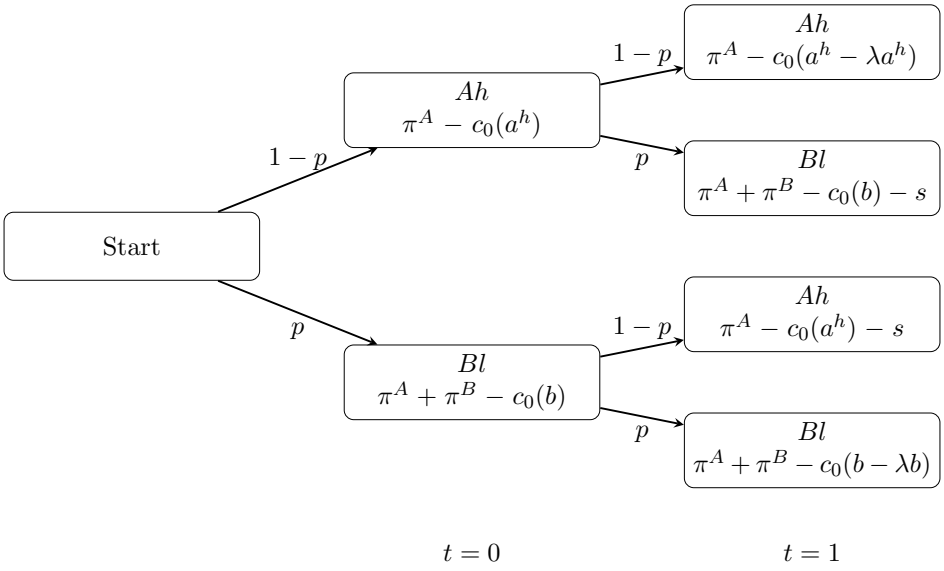
Figure 2 presents a scheme of state-contingent values in each period under the Automatic mode. The value of the Automatic mode under full learning is

$$V(\mathcal{A} \mid \lambda = 1) = 2(\pi^A + p\pi^B) - \Big(p(2-p)c_0(b) + (1-p)(1+p)c_0(a^h) + 2p(1-p)s\Big). \tag{5}$$

[13] I focus on the case of full learning since it considerably simplifies some of the expressions. In Appendix A, I derive the general results for an arbitrary value of λ.

Comparing (5) to the value of the Manual mode with full learning (4) shows that the potential extra payoff $p\pi^B$ increases the value of the Automatic mode. The costs, however, are now different due to a different task content of production, which reflects switching costs as well as reduced learning.

Figure 2: Scheme of Automatic Mode



The next step is to determine the conditions under which the agent's welfare is greater under the Automatic mode than in the Manual mode. These conditions will be equivalent to the conditions under which the agent would prefer to work under the Automatic mode, if given an option. The Automatic mode offers higher payoffs by allowing the agent to work on task B while the robot handles task Al. On the other hand, the Automatic mode induces a costly task-switching environment and slows down learning on tasks A. In a setting where the choice of technology is up to the agent (e.g., the agent owns a stake in a firm), the agent's preference will determine whether the production process will be automated (i.e., capital-intensive) or not (i.e., labor-intensive). In a setting when this choice is up to a principal, the agent's preference will determine whether the agent experiences a welfare gain or loss after the adoption of automation technology. The agent will experience a welfare gain if a principal's choice of technology coincides with the agent's preference, and a welfare loss if a principal's choice differs from the agent's preference.

The agent prefers the Automatic mode over the Manual mode if the value difference between the two modes is non-negative, $\Delta V \equiv V(\mathcal{A}) - V(\mathcal{M}) \geqslant 0$. The following result establishes the conditions under which the agent prefers the Automatic mode over the Manual mode in the case

of full learning.

Proposition 2. *Assume full learning ($\lambda = 1$). If condition*

$$\pi^B - c_0(b) \leqslant s + \frac{c_0(a^h) - c_0(a^l) - c_0(\Delta a)}{2} \tag{6}$$

holds, then there exists a unique threshold automatability $p^* \in [0, 1)$ such that the agent will prefer the Automatic mode if $p \geqslant p^*$ and the Manual mode if $p < p^*$. The threshold automatability is given by

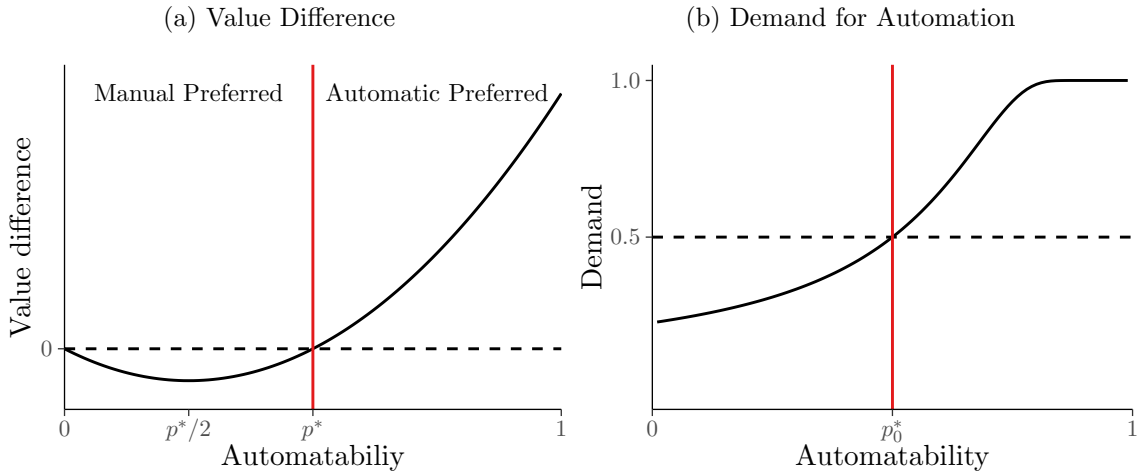
$$p^* = 1 - \frac{2\pi^B - c_0(b) + c_0(a^l)}{c_0(a^h) + c_0(b) - c_0(\Delta a) + 2s}. \tag{7}$$

If condition (6) does not hold, the agent will always prefer the Automatic mode.

This result highlights an important and intuitive interplay between the benefits and costs of imperfect complementary automation, on the one hand, and the automatability of the environment, on the other hand. Consider condition (6). The left-hand-side of this condition is the utility of the new task net of the effort cost. The right-hand-side of this condition is the sum of the switching cost and half of the term that I call the *learning cost*. Superadditivity of the cost function ensures that the learning cost is positive. The interpretation of condition (6) is that the net utility of the new task enabled by automation has to be less than the total costs (switching plus learning) imposed by automation. If this condition holds, then the automatability of the environment determines the agent's preference. If automatability is high enough, the agent will prefer the Automatic mode. If automatability is low enough, the agent would be better off under the Manual mode. However, if condition (6) does not hold and the net utility of the new task exceeds the costs of automation, then the agent's welfare will be always higher under the Automatic mode.

Even though there is a unique value at which ΔV changes its sign, the value difference is not monotonic in p. The value difference is proportional to $p(p - p^*)$, which implies that ΔV decreases in p on the interval $[0, p^*/2]$ and increases on $[p^*/2, 1]$. The value difference is negative and reaches a unique minimum at $p^*/2$. Figure 3a illustrates this point.

Figure 3: Effect of Automatability



Note: The left panel plots ΔV as a function of p. The vertical line corresponds to a threshold automatability p^*. In environments to the right of p^* the Automatic mode is preferred. In environments to the left of p^* the Manual mode is preferred. The right panel plots the demand for automation as a function of p. The vertical line corresponds to the threshold automatability for the average agent, p_0^*. In environments to the right of p_0^* the Automatic mode is chosen more than half of the time. In environments to the left of p_0^* the Automatic mode is chosen less than half of the time.

2.5 Empirical Content

The main outcome of interest—the agent's welfare—is unobservable, which makes theoretical predictions about the comparative statics of welfare untestable. To derive testable predictions, one has to focus on observable outcomes, such as choices. Suppose that we observe a sample of choices, $\{y_i\}_{i=1}^n$, made by n agents. Each agent makes a choice between either the Automatic ($y_i = 1$) or Manual ($y_i = 0$) mode. I assume that agents are heterogeneous, and their only source of heterogeneity is the switching cost. In particular, the switching cost for each agent is normally distributed with mean s_0 and standard deviation σ:

$$s_i = s_0 + \sigma\epsilon_i,$$

where ϵ_i terms are independent and identically distributed according to a standard normal distribution. An agent chooses the Automatic mode if the value difference, ΔV_i, is positive. This leads to the following result.

Proposition 3. *Assume full learning ($\lambda = 1$). The probability that agent i chooses the Automatic*

mode is

$$\mathbb{P}(y_i = 1) = \Phi\left(\frac{1}{2\sigma}\left[\frac{2\pi^B - c_0(b) + c_0(a^l)}{1-p} + c_0(\Delta a) - c_0(b) - c_0(a^h) - 2s_0\right]\right), \tag{8}$$

if $p \in (0,1)$. *For* $p \in \{0,1\}$, $\mathbb{P}(y_i = 1) = 1$.

Expression (8) can also be rewritten as

$$\Phi\left(\frac{(c_0(a^h) + c_0(b) - c_0(\Delta a) + 2s_0)(p - p_0^*)}{2\sigma(1-p)}\right),$$

where

$$p_0^* \equiv 1 - \frac{2\pi^B - c_0(b) + c_0(a^l)}{c_0(a^h) + c_0(b) - c_0(\Delta a) + 2s_0}$$

is the threshold automatability for the average agent. If $p = p_0^*$, the probability of choosing the Automatic mode will be exactly 1/2. In environments to the left (respectively, to the right) of p_0^* the probability of choosing the Automatic mode will be less (respectively, greater) than 1/2.

Let $d_{\mathcal{A}} = \mathbb{E}\, y_i = \mathbb{P}(y_i = 1)$ be the *demand for automation*, i.e., the expected proportion of agents who choose the Automatic mode. The demand for automation depends on automatability, the parameters of the switching cost distribution, learning, as well as on the payoffs and effort requirements for tasks, $d_{\mathcal{A}}(p, s_0, \sigma, \lambda, \pi^B, b, a^h, a^l)$. The following result summarizes the comparative statics for the demand. Since these results are in terms of an observable quantity—demand—they can be tested empirically.[14]

Proposition 4. *The demand for automation*

1. *Increases in automatability,* p
2. *Decreases in the mean switching cost,* s_0
3. *Increases in the standard deviation of the switching cost,* σ, *for* $p < p_0^*$ *and decreases in* σ *for* $p \geqslant p_0^*$

[14] Note, however, that since choices reflect value differences, the comparative statics for choices also reflect the comparative statics for the aggregate value difference for all agents.

4. *Increases in learning, λ, iff*

$$\frac{p}{1-p} b c_0'(b - \lambda b) \geqslant a^h c_0'(a^l - \lambda a^h)\, \mathbb{1}(a^l \geqslant \lambda a^h) + a^l c_0'(a^h - \lambda a^l) + \frac{p}{1-p} a^l c_0(a^l - \lambda a^l) \quad (9)$$

5. *Increases in the payoff for task B, π^B*

6. *Decreases in the effort requirement for task B, b*

7. *Decreases in the effort requirement for task Ah, a^h*

8. *Increases in the effort requirement for task Al, a^l iff*

$$\frac{c_0'(a^l)}{1-p} + c_0'(a^l - \lambda a^h)\, \mathbb{1}(a^l \geqslant \lambda a^h) + \frac{p}{1-p}(1-\lambda) c_0'(a^l - \lambda a^l) \geqslant \lambda c_0'(a^h - \lambda a^l). \quad (10)$$

Increasing automatability leads to an increase in the number of agents for whom automatability is greater than the threshold (7), $p \geqslant p_i^*$, raising the demand for automation. Figure 3b illustrates the effect of automatability on demand. Increasing the mean switching cost shifts the distribution of switching costs to the right. As a result of this right shift, the value of the Automatic mode decreases for all agents, which reduces the demand for automation. Increasing the variance of switching costs flattens the demand curve as a function of automatability (see Figure 3b). For $p < p_0^*$ the demand approaches 1/2 from below and for $p > p_0^*$ the demand approaches 1/2 from above as σ increases. In the extreme case of $\sigma \to \infty$ the demand will be equal to 1/2 for any $p \in (0, 1)$. The effect of learning depends, among other things, on the effort requirement for task B. Condition (9) shows that if b is high enough the effect of learning on the demand for automation will be positive. Since learning affects both types of tasks, increasing b results in benefits from learning on task B in the Automatic mode that are greater than the learning costs of task A. However, if b is not too high, e.g., $b = a^l$, condition (9) will not hold and the effect of learning on the demand will be negative.

Increasing the payoff for task B increases the value of the Automatic mode for all agents and thus raises the demand for automation. Increasing the effort requirement for task B has the opposite effect: it decreases the value of the Automatic mode and reduces the demand. Increasing the effort requirement for task Ah increases the learning costs and hence leads to a decrease in the demand for automation. The effect of the effort requirement for task Al works through learning. If $\lambda = 0$,

condition (10) will be satisfied and the effect of a^l will be positive, since in the case of no learning increasing a^l simply reduces the value of the Manual mode. On the other hand, if $\lambda \neq 0$ and a^h is high enough, condition (10) will not be satisfied and the effect of a^l will be negative.

3 Experiment

3.1 Procedures and Sample

The experiment was conducted in June–September 2019 at the ESI Lab at Chapman University. A total of 128 subjects participated in the experiment over the course of 8 sessions. Subjects in the experiment were undergraduate students at Chapman university who signed up for participation in economic experiments and accepted e-mail invitations. None of the subjects had previously experienced decision-making in an environment similar to the present experiment. Slightly more than half of the subjects, 59%, were female.[15] The average age was just above 21 years old. The majority of the subjects, 52%, identified themselves as White, while 24% self-identified as Asian, and 12% self-identified as Hispanic.

At the start of a session, an experimenter gave an overview of the experiment and answered questions. After subjects indicated that they had no outstanding questions about the general structure of the experiment, they proceeded to reading on-screen instructions that provided the details of the experiment.[16] The instructions were followed by practice rounds and the experimental task. The experiment concluded with a questionnaire and payments. The experiment was conducted on computers and programmed in z-Tree (Fischbacher, 2007).

Subjects made their decisions privately and did not communicate with each other. An independent assistant made cash payments at the end of each session. The final payoff for each subject consisted of a show-up fee, a payment for the experimental task, and a payment for belief elicitation. The payoffs ranged between \$21.25 and \$50 with an average of \$29.76, which included the \$7 show-up fee. It took subjects 55 minutes, on average, to complete the experiment.

[15] See Table D.1 in Appendix D for the detailed demographic characteristics of the sample.
[16] See Appendix B for subjects' instructions.

3.2 Experimental Task

The experimental task consisted of three phases. In the first two phases subjects experienced the *Manual* and *Automatic* modes. In the third phase subjects were given an option to choose the mode they preferred. The chosen mode is the main outcome of interest in the experiment. After subjects made their choices, they experienced the chosen mode again. The experimental task was designed to mimic the environment of the theoretical model. In particular, the task induced the salient trade-off between higher monetary benefits and higher costs in the Automatic mode. The salience of the switching and learning costs associated with the task-switching environment in the Automatic mode was the primary consideration in adopting the choice-from-experience design rather than the choice-from-description design (Hertwig and Erev, 2009).

Manual Mode

In the Manual mode, subjects were working on a real-effort task called the *Encryption* task. The Encryption task involves matching elements from one set to elements in another set according to a matching rule. This real-effort task is widely adopted in the literature (Cason, Gangadharan, and Nikiforakis, 2011; Charness, Masclet, and Villeval, 2014; Erkal, Gangadharan, and Koh, 2018). In the present implementation, subjects had to match three-digit numbers to Latin letters.[17] The matching rule (or the *key*) for the task showed the correct matches between letters and numbers and always contained 18 letter-number pairs. The key remained constant across rounds within a given phase but changed across phases. The constant key was implemented to induce the possibility of learning (Benndorf, Rau, and Sölch, 2018), since over time subjects could memorize the key and spend less effort on solving the task (Joy, Kaplan, and Fein, 2004).

In each round, subjects' had to type in the numbers that matched randomly picked letters from the key. Subjects had to match either three or six letters. Rounds in which subjects had to match three (respectively, six) letters are referred to as *easy* (respectively, *hard*) rounds.[18] The sequence of easy and hard rounds was completely random and different for each subject. A computer determined whether the round would be easy or hard before the start of each round, according to a given value

[17] See Figure C.1 in Appendix C for a sample screenshot.

[18] Experimental instructions did not use terms "easy" and "hard" in the task description to avoid priming. Instead, different rounds were referred to as having either a "3-item lock" or a "6-item lock."

of automatability. Subjects received a piece-rate of $0.25 for each correctly solved round, regardless of a round's difficulty. A round counted as correctly solved if all the letters were matched according to the key. After submitting their solutions for a given round, subjects proceeded to the next round regardless of whether the solution was correct or not.

There was no local (within a round) or global (within a phase) time constraint on task completion. Subjects could spend as much time on the task as they wished.[19] While imposing time constraints in real-effort tasks is common in the literature (Niederle and Vesterlund, 2007; Abeler et al., 2011; Gill and Prowse, 2012), time constraints were undesirable in the present design. An unconstrained design allows one to observe effort, as proxied by a response time (Spiliopoulos and Ortmann, 2018), and under certain assumptions to estimate both ability and motivation on the task (Alekseev, 2019). Observing effort was important in the present design since the theoretical model suggests that effort is one of the determinants of the demand for automation. The length of the Manual mode was determined by a fixed number of rounds, which was set to 50.

Automatic Mode

The Automatic mode differed from the Manual mode in that a computer solved easy, but not hard, rounds of the Encryption task for subjects. The computer always solved easy rounds correctly and took 10 seconds to do that. This amount of time was calibrated to match the average response time in easy rounds of the Encryption task. The theoretical model does not feature round duration, however, adopting it was necessary to induce a salient task-switching environment.

While the computer was solving easy rounds of the Encryption task, subjects were given an opportunity to work on an alternative real-effort task called the *Counting* task. The Counting task involves counting specified elements from a given set of elements and is another popular real-effort task in the literature (Abeler and Marklein, 2017; Abeler et al., 2011). Importantly, the Counting task requires skills that are different from the skills required by the Encryption task, which should induce switching costs. In the present implementation, subjects' had to count the

[19]The self-paced nature of the experiment was clearly communicated to the subjects during the instruction phase. Additional care was taken to ensure that subjects who finished early did not disturb the subjects who were still working. While it might appear that the absence of time constraints could lead to unreasonable times in the lab, this was not the case. The longest time spent in the lab by a single subject was 102 minutes, well below the two hours allocated for the experiment. Since subjects have intrinsic costs of their time, spending indefinite time in the lab is not optimal.

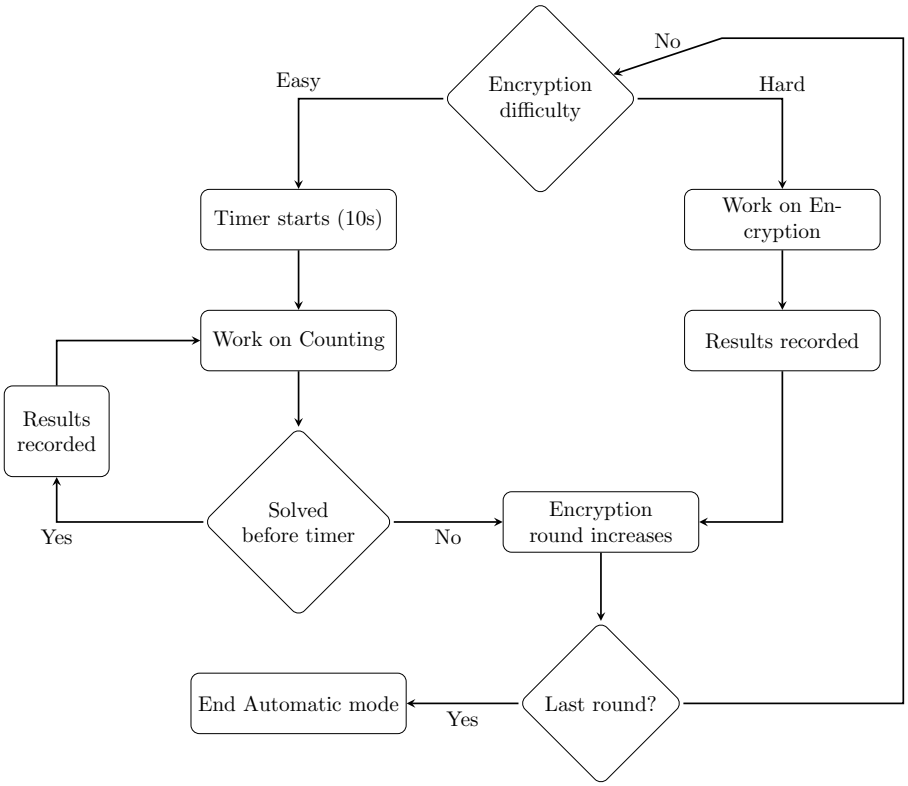
number of happy faces in a grid of happy and frowny faces.[20] The dimensions of the grid were always 4×5 elements. The number and location of happy and frowny faces was randomized in each round and was different for every subject. The number of happy faces varied between 5 and 15. The size of the grid was calibrated such that the average response time in the Counting task would match the average response time in easy rounds of the Encryption task. This was done to ensure that subjects have a chance to solve a round of a Counting task while the computer is working on an easy round of the Encryption task.

Figure 4 visualizes the timing of events in the Automatic mode. If a subject completed a round of the Counting task before the computer solved an easy round of the Encryption task, a new round of the Counting task appeared. If the computer solved an easy round of the Encryption task before the subject completed a round of the Counting task, one of the two scenarios was possible. First, the computer determined whether the next round of the Encryption task would be easy or hard. If the computer determined the next round to be easy, the computer silently proceeded to working on that round, and the subject continued working on the Counting task for at least another 10-second period. If the computer determined the next round to be hard, the computer interrupted the subject's work on the Counting task and presented the subject with a hard round of the Encryption task.[21] After the subject completed a hard round of the Encryption task, the computer determined the next round's difficulty, and so on.

As in the Manual mode, the Encryption task did not have local or global time constraints for subjects, but had a fixed number of rounds, 50. Unlike in the Encryption task, the number of rounds in the Counting task was unlimited, though there was a certain time constraint. Subjects could complete as many rounds of the Counting task as possible during the time a computer was solving easy rounds of the Encryption task. The Automatic mode finished when the last round of the Encryption task was completed either by a subject or a computer. Subjects received a piece-rate of \$0.25 for each correctly solved round of the Counting task, as well as for each correctly solved round of the Encryption task, *including* the rounds completed by the computer. This implementation of the higher monetary benefits in the Automatic mode relative to the Manual mode was chosen because it makes transparent for the subjects that they can *never earn less* money in the Automatic

[20] See Figures C.2 in Appendix C for a sample screenshot.
[21] See Figures C.3 in Appendix C for a sample screenshot.

Figure 4: Structure of Automatic Mode



mode than in the Manual mode.[22]

3.3 Feedback and Incentive Mechanism

Subjects did not receive feedback on their performance neither after each round nor after each phase. Subjects received the feedback, which included the number of correctly solved rounds in both tasks and total earnings in each phase, only at the end of the experiment. The after-phase feedback was omitted for two reasons. The first reason was to have an opportunity to elicit subjects' beliefs about their overall (i.e., in all three phases) performance in both tasks. These beliefs could only be elicited after the last phase. The second reason was to enhance the salience of the task-switching environment. While the monetary advantage of the Automatic mode was clearly explained to subjects during instruction, the absence of after-phase feedback prompted subjects to rely on how

[22] An alternative implementation could be, e.g., to pay a higher piece rate for the Counting task and not pay for the work done by the computer. However, in this implementation the monetary advantage of the Automatic mode would be unclear, since it would depend entirely on subjects' actual performance in the Counting task. In the preferred implementation, the Automatic mode always yields higher earnings regardless of subjects' performance on the Counting task. The exact magnitude of the extra earnings does depend on performance.

each mode felt rather than on the exact information about earnings.

The after-round feedback was omitted to ensure the independence of rounds. Providing after-round feedback could lead to situations when learning about the performance in one round affects subsequent performance. Situations like these would complicate the evaluation of subjects' actual and perceived performance.

The payoff for the experimental task was determined by a computer that randomly selected one phase out of three. This random selection mechanism ensures that subjects' behavior in each phase is incentive compatible in terms of performance on the real-effort tasks and the choice of a preferred mode (Azrieli, Chambers, and Healy, 2018). The mechanism is frequently used in experiments where subjects experience different institutions and then make a choice for the preferred one (Niederle and Vesterlund, 2007).

3.4 Treatments, Hypotheses, and Power

The experiment used a 2×2 full-factorial between-subject design. The first treatment arm was automatability, which determined the probability of an easy Encryption task in each round. An assigned value of automatability remained constant in every phase of the experimental task. The automatability had two levels: 20% (low) and 80% (high). Proposition 4 (part 1) implies the following testable hypothesis.

Hypothesis 1. *The demand for the Automatic mode should be higher under high automatability than under low automatability.*

The second treatment arm was the order of modes that subjects experienced before making their choices in the third phase. This treatment arm was introduced to explore a potential order effect. The order treatment arm had two levels: AM (Automatic mode experienced first) and MA (Manual mode experienced first). Theoretical model does not make any predictions about the order effect, however, there could be behavioral mechanisms at play that would lead to such an effect.

The treatment assignment followed a completely randomized design (Selten, Friedman, and Cassar, 2004). The computer randomized treatment assignment within a session on a subject level. Table 1 shows the balance of subjects across the treatment cells. The target sample sizes of approximately 63 subjects per value of a treatment arm were chosen to yield an 80% power to

detect a medium effect size (Cohen's $h = 0.5$) at a 5% significance level. The actual sample sizes differ slightly from those targets due to the completely random nature of treatment assignment.

Table 1: Treatment Balance

	Order		
Automatability	AM	MA	Total
0.2	35	29	64
0.8	31	33	64
Total	66	62	128

Note: The table shows the number of subjects in each treatment cell, as well as the total number of subjects for each value of each treatment arm.

3.5 Questionnaire

After completing the experimental task, subjects participated in a questionnaire that contained additional instruments aimed at exploring the determinants of choices beyond treatment variables. I discuss these additional questions and the motivation for their inclusion below.

Questions about Choices

The first question asked subjects to explain the reasons behind their choices.[23] The question was administered at the start of the questionnaire. The answers to this question help evaluate how well subjects' own reasons match *a priori* motivations suggested by the theoretical model. Subjects gave their answers as a free-form text, which was classified by an independent research assistant.[24]

The second question asked subjects whether they would revise their choices, if given an option to do so. The question was administered after subjects received the feedback on their performance and earnings, but before the final payoff was randomly selected. Subjects' answers did not affect payoffs, which was clearly explained. The answers to this question help assess the impact of feedback on choices and provide insights into whether subjects' choices were consistent.[25]

[23] Textual analysis of reasons is an example of process data that gains popularity in economics (Cooper, Krajbich, and Noussair, 2019). See Capra (2019) for examples and a discussion of protocol analysis in economics.

[24] See Appendix E for the details of the classification procedure.

[25] The answer to this hypothetical question, however, can either under- or overestimate the actual choice consistency. Even if someone feels like they made a mistake, they might want stick to the original answer to preserve a good self-image of themselves (Ewers and Zimmermann, 2015). This effect would bias the estimate of choice consistency

Task Enjoyment and Beliefs

The enjoyment question asked subjects to rate their subjective enjoyment of the Encryption and Counting tasks. The answers were given on a five-point Likert scale. Task enjoyment can be thought of as a proxy for the intrinsic utility of working on a task. Since the intrinsic utility of a task enters the overall value of a mode and can affect choice, it is important to control for it.

Belief questions asked subjects to guess their perceived accuracy (in percentage points) and mean response time (in seconds) on the Encryption and Counting tasks. Since subjects made their choices prior to seeing feedback, it was likely that beliefs about performance affected subjects' choices. Questions about the Encryption task were split by difficulty. Subjects had to consider only the rounds they completed by themselves and the rounds in which they were not interrupted by the computer. The reports were incentivized using a quadratic scoring rule, presented as a schedule for clarity. The payoff for each belief question was determined as $\$1 \times \big(1 - 0.0025(\text{actual value} - \text{report})^2\big)$, so that the maximum a subject could earn for each of the six belief questions was \$1. Belief questions were not announced prior to the experimental task, hence it was impossible for subjects to manipulate their performance to make sharp predictions during the belief elicitation.

Demographic Questions

This set of questions included the standard items: gender, age, ethnicity, marital status, year in school, major, GPA, weekly expenditures, work on campus (wage and hours), family income level, and religious affiliation. Gender was a particularly important characteristic to consider given the prominence of a gender wage-gap in policy discussions. Studying the effect of gender on choices helps address the question of whether automation would help to close this gap or make it wider.

Cognitive Reflection Test (CRT)

The CRT (Frederick, 2005) contained three standard questions, as well as a question about whether subjects saw the CRT before. The CRT score shows subjects' reliance on slow versus fast systems in their decision-making and is related to cognitive flexibility. Since CRT is a measure of general

upwards. It is also possible that someone reports that they wished they could revise their choice, but in reality they would choose the same thing, an effect similar to time-inconsistency. This effect would bias the estimate of choice consistency downwards.

intelligence, it could interact with subjects' perceptions of benefits and costs of a given mode and ultimately affect their choices.

Big Five

The Big Five questionnaire used a 50-question NEO-PI-R Domain inventory by Costa and McCrae (1992) made available via International Personality Item Pool (ipip.ori.org).[26] The inventory assesses the five personality domains (10 questions per domain): Openness to Experience, Conscientiousness, Extroversion, Agreeableness, and Neuroticism. The domain-specific questions were presented in a random order with plus- and minus-keyed items alternating.

Big Five questions help explore whether certain types of personalities have a stronger preference for either of the two modes. For example, people with high Openness to Experience are generally willing to try out new things. If the Automatic mode is perceived as a novelty, subjects who score highly on this trait would be more likely to choose the Automatic mode. For another example, one might expect subjects who score highly on Neuroticism to be less likely to choose the Automatic mode because they poorly handle the stress associated with task switching (Afshar et al., 2015).

Including these questions was important for two reasons. First, recent research in economics suggests that personality traits play a role in subjects' behavior in a labor market context (Filiz-Ozbay et al., 2018). Second, companies often use personality tests as a part of their recruitment practices. The existence of a relationship between personality and preferences for automation could, therefore, inform the corporate policies on the adoption of automation technologies.

Risk and Time Preferences

The risk and time elicitation followed the design proposed by Falk et al. (2016) and recently used in Falk et al. (2018). There were two parts for each type of preferences: a qualitative and a quantitative part. The qualitative part asked subjects about their general tendency to either take risks or delay gratification on a 10-point Likert scale. The quantitative part asked subjects to state their hypothetical preference for either a sure amount versus a lottery (for a risk preference) or a payment today versus a payment in the future (for a time preference). The quantitative part was presented in an unfolding brackets format in which the next question depends on the previous

[26] Borghans et al. (2008) provide an excellent discussion of the Big Five and the meaning of the domains.

answer. The answers to the qualitative and quantitative parts were weighted using the weights in Falk et al. (2018)[Appendix 1.I].

Attitudes towards risk and time potentially could be important determinants of subjects' choices in a labor context (Niederle and Vesterlund, 2007; Corgnet and Hernán-González, 2019). While both the Automatic and Manual modes involve uncertainty about an upcoming round, the uncertainty embedded in the Automatic mode is arguably larger. Patience can affect choice since the Automatic mode can be viewed as entailing a costly investment today that bears productivity gains in the future.

3.6 Performance Measures

I compute several measures of performance on the Encryption and Counting tasks. I compute the mean response time (RT) (average time needed to complete a round) and accuracy (the fraction of rounds solved correctly) for each task using the data from the first two phases. In addition to the observable performance measures, I estimate the structural measures of ability and motivation using the methodology developed in Alekseev (2019). Ability is a measure of efficiency that represents how quickly subjects can correctly solve a task, while motivation is a measure of intrinsic utility that represents how much subjects care about correctly solving a task. Given that these structural measures are simple non-linear transformations of mean RT and accuracy, I also derive the *perceived* ability and motivation using subjects' beliefs about mean RT and accuracy.

4 Results

4.1 Summary

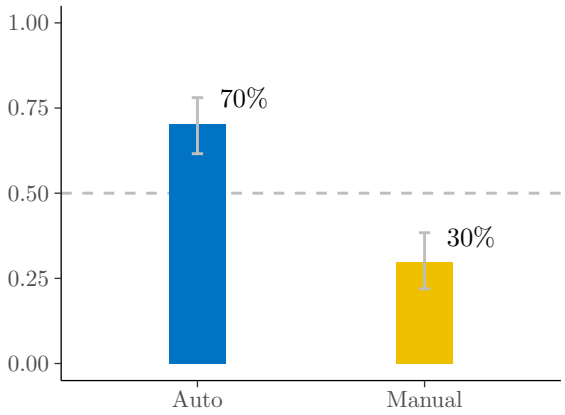
I begin the analysis by looking at the unconditional distribution of choices in the sample. Figure 5 shows that the majority of subjects, 70%, chose the Automatic mode. Yet a non-trivial proportion of subjects, 30%, chose the Manual mode. The frequency of the Automatic choice is significantly greater than the frequency of the Manual choice ($p-$value < 0.001, Exact binomial test).[27]

The observed variation in subjects' choices is inconsistent with the model in which subjects only care about money. Recall that the Automatic mode always dominates the Manual mode in terms

[27] All the statistical tests reported in the paper are two-sided, unless otherwise noted.

of earnings. If subjects only cared about money, everyone would have chosen the Automatic mode, which is clearly not the case. The frequency of the Manual choice is also too high to be entirely explained by choice errors on the part of subjects.[28] Therefore, subjects must have considered the costs associated with the Automatic mode, and for some subjects, these costs outweighed the benefits of automation. The subsequent sections will explore the sources of the observed variation in subjects' choices.

Figure 5: Unconditional Distribution of Choices



Note: The figure shows the frequencies of choices of each mode in the entire sample. The error bars show the 95% confidence interval.

4.2 Treatment Effects

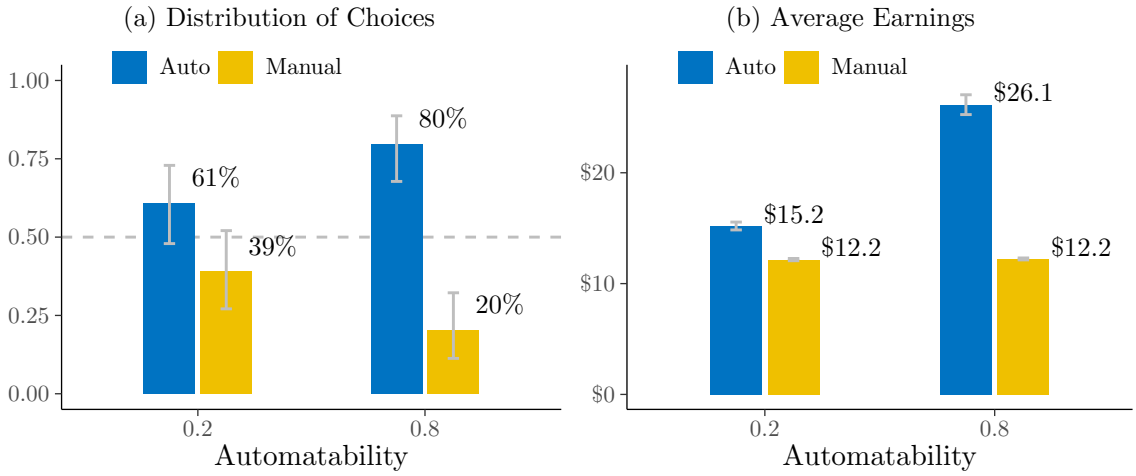
Automatability

The first potential candidate to explain the observed variation in subjects' choices is automatability. Figure 6a shows the distribution of subjects' choices conditional on this treatment variable. Under low automatability, the frequency of the Automatic choice is 61%. Under high automatability, the frequency of the Automatic choice is 80%. This implies a positive average treatment effect (ATE) of automatability on the Automatic choice of 18.75 percentage points (ppts) (Cohen's $h = 0.42$, medium effect size). The treatment effect is statistically significant ($p-$value $= 0.033$, Fisher's Exact Test for Count Data).

Result 1. *The demand for the Automatic mode increases in automatability.*

[28] I elaborate on the point about potential choice errors in Section 4.4.

Figure 6: Treatment Effect of Automatability



Note: The figure shows the distribution of choices (left panel), as well as average earnings (right panel), conditional on automatability. The error bars show the 95% confidence intervals.

The direction of the treatment effect is in line with the theoretical prediction and thus supports Hypothesis 1. However, two quantitative features of Figure 6a deserve a closer inspection. First, not everyone chose the Automatic mode under high automatability. The proportion of subjects who did not choose the Automatic mode is, in fact, 20%. Such a high demand for the Manual mode under high automatability is surprising, because subjects could earn substantially more money in the Automatic mode than in the Manual mode. Figure 6b shows that by choosing the Automatic mode under high automatability subjects could have more than doubled their earnings relative to the Manual mode.[29] One potential explanation for this result is that the huge monetary advantage of the Automatic mode under high automatability was not salient enough for some subjects. Second, the Automatic mode remains the dominant choice even under low automatability. This is surprising because the simulation on Figure 3b suggests that the frequency of the Automatic choice should be less than 50% when automatability is low. The data, on the other hand, clearly reject the hypothesis that the frequency of the Automatic choice under low automatability is less than 50% ($p-$value $= 0.97$, Exact binomial test). I explore these issues in more detail in Section 4.4.

[29]Subjects could never earn less money in the Automatic mode than in the Manual mode by design.

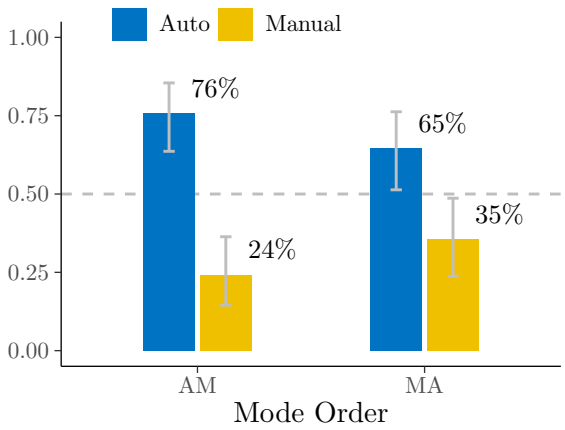
Order Effect

While the theoretical model does not make any predictions about the order effect, it is conceivable that the mode order could affect subjects' choices. For example, experiencing a given mode first could create an anchor for this mode and thus make it more appealing. Under this hypothesis subjects would tend to choose the Automatic mode at a higher rate in the *AM* order (Automatic mode first) than in the *MA* order (Manual mode first). An alternative hypothesis could be that subjects anchor on the mode they experience last. Under this hypothesis, subjects would tend to choose the Automatic mode at a lower rate in the *AM* order than in the *MA* order. It is also possible for both effects to be present in the sample. This would make the order effect less pronounced.

Figure 7 shows the distribution of subjects' choices conditional on mode order. Under the *AM* order, the frequency of the Automatic choice is 76%. Under the *MA* order, the frequency of the Automatic choice is 65%. This implies a positive ATE of experiencing the Automatic mode first on the Automatic choice of 11.24 ppts (Cohen's $h = 0.25$, small effect size). The treatment effect, however, is not statistically significant ($p-$value $= 0.18$, Fisher's Exact Test for Count Data). This suggests that either the order effect was weak in the sample or that the two opposing order effects were present and canceled each other.

Result 2. *The demand for the Automatic mode is not significantly affected by the mode order.*

Figure 7: Distribution of Choices by Mode Order



Note: The figure shows the distribution of choices conditional on the order in which modes were experienced. *AM* denotes that the Automatic mode was experienced first, *MA* denotes that the Manual mode was experienced first. The error bars show the 95% confidence intervals.

Regression

I conclude the discussion of treatments effects with a regression analysis. Table 2 shows the results of a logit regression with both treatment variables, as well as their interaction, included as regressors.[30] The regression results corroborate the results obtained from non-parametric tests. Increasing automatability has a positive effect on the probability of the Automatic choice. Experiencing the Automatic mode last (*MA* order), on the other hand, appears to slightly reduce the probability of the Automatic choice, however, the treatment effect is not statistically significant. There is no evidence of the interaction between the two treatment variables.

Table 2: Logit Regression Results

Variable	Coefficient	SE	Statistic	p-value
Constant	0.651	0.434	1.499	0.134
Automatability	1.259	0.549	2.294	0.022
Order	-0.443	0.358	-1.237	0.216
Automatability × Order	-0.486	0.619	-0.784	0.433

Note: The table shows the estimation results from a logit model. The dependent variable is whether a subject chose the Automatic mode. Automatability variable refers to the case when automatability is 0.8. Order variable refers to the case when order is *MA*. The standard errors are heteroskedasticity-robust (HC1) and clustered on the session level.

4.3 Exploratory Analysis

Having established the causal effect of automatability on the propensity to choose the Automatic mode, I turn to exploring which subjects' characteristics could further explain the observed variation in choices. The nature of the dataset suggests focusing on a predictive *out-of-sample* analysis as a opposed to a more traditional explanatory *within-sample* analysis, for at least two reasons. The first reason is a statistical one. The large number of potential predictors raises the issue of overfitting, which occurs when a model successfully predicts the outcomes within sample but fails to predict the outcomes out of sample. The second reason is a practical one. The predictive performance of a model is more relevant, than explanatory performance, for designing policies and evaluating their potential effects.

[30] I estimate a long model, i.e., with the interaction term, following Muralidharan, Romero, and Wüthrich (2019). Since the present experiment uses a full factorial design, ignoring the interaction term would result in incorrect inferences.

To conduct variable selection in a disciplined manner, I use the Lasso estimator, which is a standard tool in machine learning that is gaining popularity in economics (Varian, 2014; Mullainathan and Spiess, 2017). The Lasso is a linear model that penalizes for the absolute magnitude of coefficients, shrinking some coefficients to exactly zero. While certainly being not as flexible as other machine learning methods, such as, e.g., Random Forest, the Lasso is appealing for its tractability. The Lasso serves as a convenient bridge between the modern machine learning techniques and a standard economic analysis of effect sizes.

Estimation Strategy

I use a nested cross-validation (CV) algorithm (Varma and Simon, 2006; Krstajic et al., 2014) for analysis. The algorithm consists of two loops: an outer loop and an inner loop. The outer loop performs a repeated CV to obtain an unbiased estimate a model's out-of-sample performance.[31] All of the observations are split into k^{outer} equally-sized subsets (folds). One of the folds is held out as a *test* set used to evaluate the model's out-of-sample performance, while the remaining folds are used as a *training* set for model estimation and tuning. The process is repeated until all of the k^{outer} folds have been used as test sets. The splitting into folds is repeated n^{outer} times to further reduce the bias introduced by random splitting.

The inner loop applies to each of the training sets drawn at the outer loop. The Lasso has a tuning parameter that determines the magnitude of the penalty. This parameter cannot be estimated from the data, hence a repeated CV is used to *tune* it. The training set is split into k^{inner} folds. One of the folds is held out as an *assessment* set, while the remaining folds are used as an *analysis* set for model estimation. The model is estimated for a grid of values of the penalty parameter, and the model's performance on the assessment set is estimated for each grid value. The process is repeated until all of the k^{inner} folds have been used as analysis sets. As in the outer loop, the splitting into folds is repeated n^{inner} times to further reduce the bias. The model's performance across all the assessment sets is averaged for each grid value, and the value of the tuning parameter that maximizes the performance is used as the best value on a given test set.

Given that the distribution of outcomes in my dataset is highly skewed towards the Automatic

[31] The bias in estimating the out-of-sample performance can become a real issue given the small number of observations in the dataset.

mode, I use stratification when creating folds to preserve the distribution of the outcomes in each fold (Kohavi, 1995). I use the area under the ROC curve (AUC) as a measure of a model's performance (DeLong, DeLong, and Clarke-Pearson, 1988; Bradley, 1997). AUC is the standard performance measure in machine learning when the outcome variable is categorical. I set $k^{\text{outer}} = 4$, $k^{\text{inner}} = 10$, and $n^{\text{inner}} = n^{\text{outer}} = 16$, which results in a total of 10240 models being estimated.

The nested CV produces a distribution of performance estimates, as well as the distribution of the best tuning parameters for the Lasso. I use the median value of the best tuning parameters to estimate the Lasso on the full dataset and obtain the selected predictors.[32] I then use the selected predictors to estimate a standard logit model to get the effect sizes for the predictors.

Results

I begin by addressing the issue of overfitting formally and compare the out-of-sample performance of the Lasso and the logit models. The logit models use all of the available predictors. The mean AUC for the Lasso model is 0.69, while the mean AUC for the logit model is 0.62.[33] The Lasso performs better than the logit in 80% of the cases, and the average relative gain in the AUC is 10%. The difference in the performance between the two models is statistically significant ($p-$value < 0.001, Wilcoxon signed rank test with continuity correction).

Figure 8 shows the estimation results from the logit model that only uses the predictors selected by the Lasso. The subset of selected predictors, which has only 5 variables, is tiny compared to the overall set of potential predictors, which has 48 variables. Among the selected predictors, all but one are related to the Counting task. The only predictor not related to the Counting task is the CRT score. The Counting task predictors include a measure of subjective task enjoyment, as well as performance measures. Both subjective (beliefs) and objective (actual) performance measures survive the selection process. The selected performance measures include the mean response time (RT) and motivation. None of the measures of accuracy or ability survive the selection, which suggests that when choosing a mode, subjects tend to care more about how quickly they can perform a task rather than how well the can perform it. It is also important to note that among the

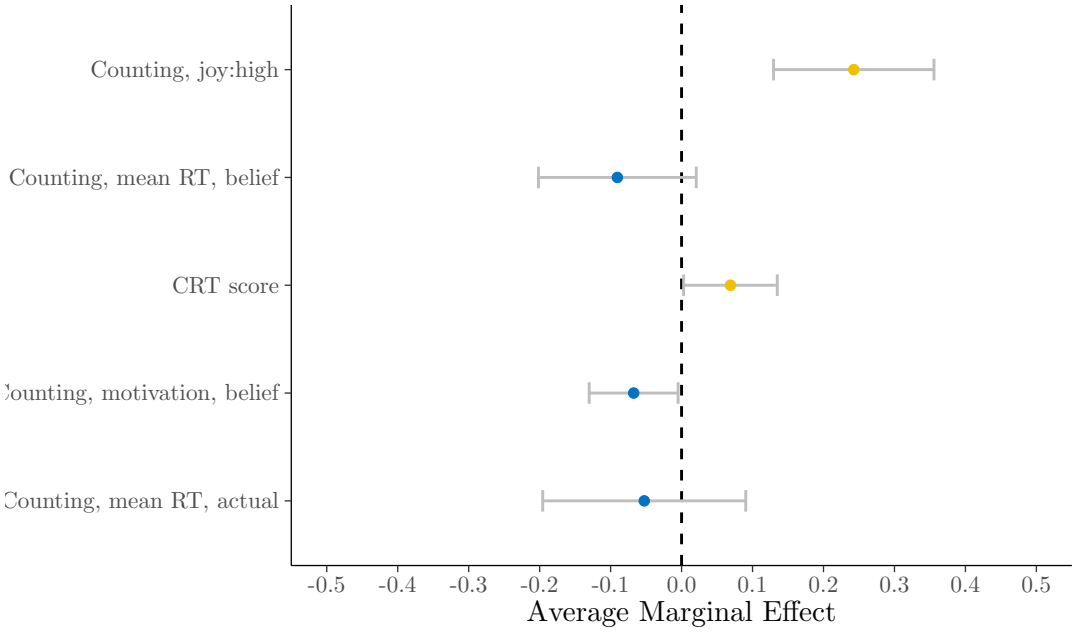
[32] Note that I do not use the out-of-sample performance on the test sets to pick the best-of-the-best tuning parameter. Such a procedure would lead to overfitting.

[33] A model with just a constant term has an AUC of 0.5, while the perfectly predictive model would have an AUC of 1.

predictors that *did not* survive the selection are all of the demographic, preferences, and personality variables.

Result 3. *The most important predictors of demand for the Automatic mode are Counting task enjoyment and performance, as well as the CRT score.*

Figure 8: Average Marginal Effects



Note: The figure plots the estimated AMEs of the variables selected by the Lasso from a Logit regression. Continuous variables are standardized. The horizontal bars are 95% confidence intervals (heteroskedasticity-robust (HC1), clustered at the session level). The dependent variable is whether a subject chose the Automatic mode.

Turning to effect sizes, we see that subjective task enjoyment has a particularly large and positive effect on the propensity to choose the Automatic mode. This effect is intuitive and consistent with the model's predictions. A higher intrinsic utility of the Counting task, as proxied by task enjoyment, leads to a higher payoff π^B, which by Proposition 4 (part 4) should increase the demand for the Automatic mode. The mean RT on the Counting task can theoretically affect the demand for the Automatic mode via two channels. First, higher mean RT implies a higher effort requirement b, which by Proposition 4 (part 5) should decrease the demand. Second, a subject with high mean RT cannot complete as many trials of the Counting task as a subject with low mean RT, hence higher mean RT implies lower payoff π^B and thus should decrease the demand. The negative

effect of perceived motivation is somewhat counter-intuitive, given that higher motivation should theoretically increase the payoff π^B. However, higher motivation would also result in higher mean RT, which has a negative effect on the demand. The overall negative effect of motivation suggests that the latter negative effect (via higher mean RT) dominated the former positive effect (via higher payoff). The effect of the CRT score can be understood through the lens of studies in psychology on cognitive flexibility (Martin and Rubin, 1995; Chung, Su, and Su, 2012). If the CRT score is negatively correlated with the switching costs, then Proposition 4 (part 3) would, indeed, imply the positive effect of the CRT score on the demand for the Automatic mode.

4.4 Choice Reasons

In order to obtain additional insights into subjects' behavior, I elicit subjects' responses about the reasons behind their choices. Figure 9 shows that the dominant reason for choosing the Automatic mode was its earnings advantage. This reason was given by subjects 48% of the time (each subject could give multiple reasons). Among other popular reasons were that the Automatic mode felt faster, stated 21% of the time, and a closely related reason that the Automatic mode felt less monotonous, stated 15% of the time.

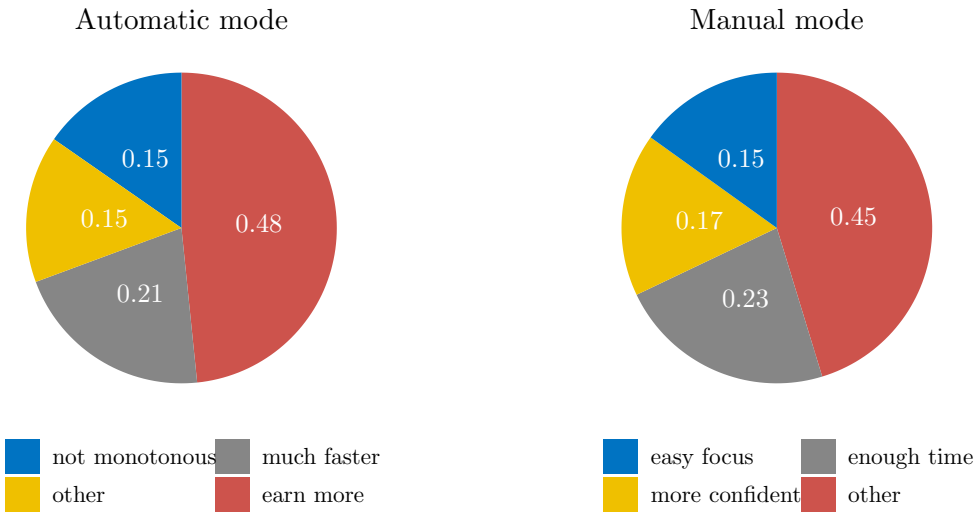
The reasons for choosing the Manual mode were more idiosyncratic, as can be judged by the size of the "other" category. One of the common reasons for choosing the Manual mode was the desire to have enough time to work on the task. This reason was given 23% of the time. Other reasons included a greater feeling of confidence, stated 17% of the time, and that subjects found it easier to focus, stated 15% of the time.

Subjects' responses leave the impression that subjects made well-motivated choices.[34] Some of the revealed reasons behind subjects' choices, such as monetary benefits, are explicitly captured by the theoretical model. Some other prominent reasons, however, are not explicitly captured by the model and could not be recovered using choice data alone. Consider, for example, the two reasons for choosing the Automatic mode classified as "not monotonous" and "much faster."[35] A potential

[34] One might argue that some of the stated motives could not be that strong to reasonably justify a given choice, but this would be a comment on whether subjects' preferences are reasonable rather than on whether choices reflected preferences, whatever these preference were.

[35] The "much faster" reason can be related to the "earn more" reason, since being able to complete the task faster results in higher earnings. However, it appears that subjects perceive these two reasons as distinct, as is evident from their responses, and therefore, these two reasons represent different mechanisms.

Figure 9: Choice Reasons



Note: The figure presents the breakdown of choice reasons by categories for each mode choice. Subjects could give multiple reasons for their choices.

interpretation of these reasons is that the task switching environment introduced by the Automatic mode had unexpected effects on some subjects. Instead of increasing the *disutility* of those subjects via higher switching costs, task switching increased their *utility* via a less monotonous work process. This effect is consistent with the findings in the organizational psychology on the positive effect of task variety for some workers (Zaniboni, Truxillo, and Fraccaroli, 2013). A less monotonous work process is also often perceived as going fast (Loukidou, Loan-Clarke, and Daniels, 2009). The model can incorporate such effects by allowing the switching costs to be negative for some subjects, i.e., in these cases switching costs would turn into switching benefits.

The reasons for choosing the Manual mode are harder to map directly into the motivations in the model. In part this is due to the large number of idiosyncratic reasons that fall under the "other" category. The "easy focus" reason suggests that some subjects find it easier to focus on tasks when tasks are of a single type. Easier focus can be interpreted as less difficulty with completing the task and hence less effort. This effect is similar to the learning process assumed in the model. The "more confident" reason can be related to the perceived accuracy of subjects' performance on the tasks. If task switching reduces subjects' confidence in the correctness of their answers, subjects would need to put in more effort to make sure the tasks are solved correctly. This effect is also reminiscent of the learning process. It would, however, be more accurately represented

by a "negative learning" where the effort requirement increases whenever task switching occurs. The "enough time" reason is related to the sense of time pressure in the Automatic mode. Even though the model does not explicitly incorporate time pressure, the psychological stress induced by time pressure can be captured in switching costs.

Result 4. *The dominant stated reason for choosing the Automatic mode is its monetary advantage. The majority of the stated reasons for choosing the Manual mode are idiosyncratic, however, many of the reasons are related to learning.*

4.5 Choice Consistency

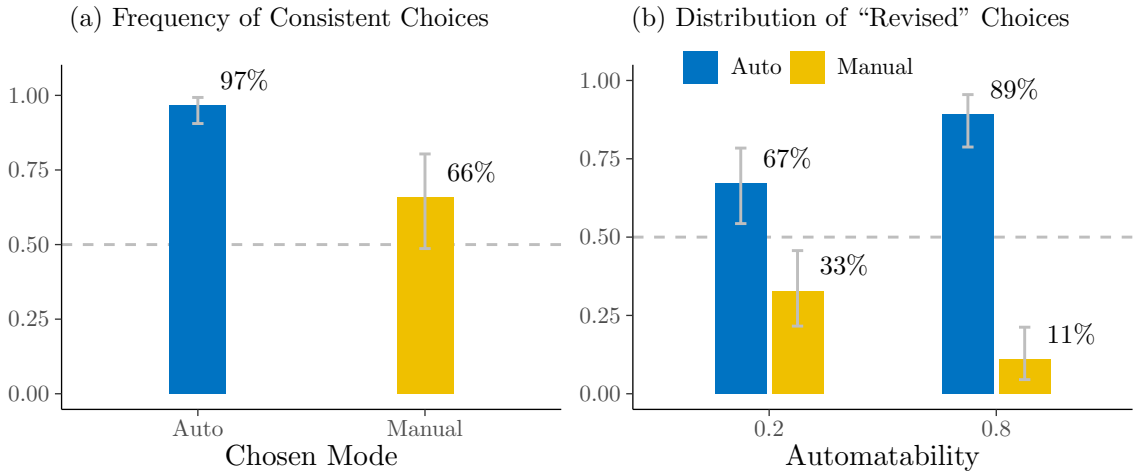
Part of the reason why we observe lower-than-expected demand for automation in the high automatability treatment could lie in subjects' misperception of costs and benefits of the two modes. To explore this possibility, I turn to the analysis of subjects' choice consistency. I elicit choice consistency by computing the proportion of subjects who answered that they would not revise their choices. Recall that the question about whether a subject wished to revise his or her choice, if given an option, was administered after learning the feedback on earnings in each mode. A subject who initially underappreciated the magnitude of potential earnings in the Automatic mode could update his or her beliefs accordingly after learning this feedback.

The majority of subjects, 88%, made consistent choices. Figure 10a shows, however, that choice consistency varies dramatically by the chosen mode. Among subjects who chose the Automatic mode, 97% made consistent choices. In contrast, among subjects who chose the Manual mode, only 66% did so. The difference in the choice consistency between the chosen modes is large (30.88 ppts) and statistically significant ($p-$value < 0.001, Fisher's Exact Test for Count Data).

Result 5. *The average choice consistency is high. Choice consistency is substantially lower for subjects who chose the Manual mode than for those who chose the Automatic mode.*

Imperfect choice consistency raises the question about what would happen to the treatment effect of automatability if subjects were allowed to revise their choices. Figure 10b addresses this question by showing the distributions of "revised" choices conditional on automatability. The demand for automation would go up under both values of automatability, however, the treatment

Figure 10: Choice Consistency and "Revised" Treatment Effect



Note: The figure shows the frequency of consistent choices by the chosen mode (left panel), as well as the distribution of "revised" choices conditional on automatability (right panel). The error bars show the 95% confidence intervals.

effect would be robust to this change.[36] In fact, the treatment effect would increase up to 21.88 ppts (Cohen's $h = 0.55$, medium effect size) and gain in statistical significance ($p-$value $= 0.005$, Fisher's Exact Test for Count Data).

5 Conclusion

In this paper, I have presented evidence that the welfare effect of complementary automation on workers is not always positive compared to what is suggested by the current literature. Complementary automation does enable new, better tasks for workers and increase their productivity and wages. However, new technologies are often imperfect, and these imperfections can impose non-pecuniary costs on workers via a behavioral channel. The adoption of "so-so technologies" (Acemoglu and Restrepo, 2019), while productivity-enhancing from the viewpoint of management, can, in reality, lead to net welfare losses for the workers who have to deal with such technologies on a day-to-day basis, as the recent Walmart example illustrates.

I formalize this intuition in a model, inspired by the task-based framework of Acemoglu and Restrepo (2018b). I show that the welfare effect of complementary automation is determined by the

[36] It is interesting to note that, even allowing for the possibility to revise a choice, the demand for automation under high automatability would be less than 100%.

trade-off between the benefits and costs of automation. The benefits of automation in this setting stem from new, better tasks enabled by the technology. The costs stem from the imperfection of the technology that creates a task-switching environment. I use the model to identify several key factors that can tip the balance of benefits and costs. The automatability of the production process emerges as a particularly important factor. For a single agent, I show that the net welfare effect of automation is determined by a threshold value of automatability. If the actual automatability of the production process exceeds this threshold, the agent will always be better off with automation, and *vice versa*. I also derive testable predictions for the demand for automation observed in a sample of heterogeneous agents.

I test the model's predictions and generate rich empirical data using an experiment. The experiment elicits subjects' revealed preference for automation using a choice-from-experience design (Niederle and Vesterlund, 2007; Hertwig and Erev, 2009). I vary the automatability of the production process between subjects and observe how this exogenous variation affects the demand for automation. In line with the model's predictions, increasing automatability does lead to a higher demand for automation. I explore other potentially important predictors of the demand for automation using a machine learning analysis. Only a few such predictors emerge from a large set of potential predictors, including demographic characteristics, risk and time preferences, personality traits, and task performance. In particular, measures of task enjoyment, performance, and cognitive flexibility tend to be the strongest predictors of the demand for automation. The textual analysis of subjects' reasons for their choices yields additional insights into the determinants of the demand. The analysis reveals that some subjects enjoy a task-switching environment and experience a positive effect of task variety instead of an adverse effect of task switching.

The theoretical and empirical analysis in this paper generates two new findings about the potential effects of complementary automation on workers. First, the net welfare effect of complementary automation depends on the parameters of the environment, such as the automatability of the production process. In some environments, the net welfare effect of automation can become negative, even when automation does generate productivity gains. The existence of behavioral effects that arise when a worker interacts with the technology is what drives this result. Second, there is an apparent heterogeneity of workers in terms of how much they benefit from automation. For example, subjects with higher cognitive flexibility are more likely to go with automation. Since cognitive

flexibility is associated with IQ, and higher IQ is associated with higher earnings (Acemoglu and Autor, 2011), complementary automation, even when its effect is positive, can amplify welfare inequality among workers.

These findings suggest that the future discussion of the potential effects of automation should acknowledge behavioral factors. Acknowledging these factors will help design better policies that alleviate the potential adverse effects of imperfect automation. The introduced theoretical framework offers several distinct ways to address these adverse effects. For example, increasing the monetary payoffs for the new tasks enabled by automation or increasing their non-monetary attractiveness can boost the net welfare effect of automation. Alternatively, one could focus on improving the automatability of the environment or on how workers interact with technology. Designing automation technology in a way that reduces the switching costs for workers could be an essential tool to increase the net welfare effect of automation. Comparing different compensation methods and evaluating their effectiveness thus appears to be a promising direction for further research. Finally, given the unequal distribution of the net welfare effects of automation among workers, it would be crucial to identify the groups of workers who would benefit most from these compensation methods.

References

Abeler, J., A. Falk, L. Goette, and D. Huffman. 2011. "Reference Points and Effort Provision." *American Economic Review* 101:470–92.

Abeler, J., and F. Marklein. 2017. "Fungibility, Labels, and Consumption." *Journal of the European Economic Association* 15:99–127.

Acemoglu, D., and D. Autor. 2011. "Skills, Tasks and Technologies: Implications for Employment and Earnings." In D. Card and O. Ashenfelter, eds. *Handbook of Labor Economics*. Elsevier, vol. 4, chap. 12, pp. 1043–1171.

Acemoglu, D., and P. Restrepo. 2018a. "Artificial Intelligence, Automation and Work." Working Paper No. 24196, National Bureau of Economic Research, January.

—. 2019. "Automation and New Tasks: How Technology Displaces and Reinstates Labor." *Journal of Economic Perspectives* 33(2):3–30.

—. 2018b. "The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment." *American Economic Review* 108:1488–1542.

—. 2020. "Robots and Jobs: Evidence from US Labor Markets." *Journal of Political Economy* 128:2188–2244.

Afshar, H., H.R. Roohafza, A.H. Keshteli, M. Mazaheri, A. Feizi, and P. Adibi. 2015. "The Association of Personality Traits and Coping Styles According to Stress Level." *Journal of Research in Medical Sciences* 20:353–358.

Aghion, P., B.F. Jones, and C.I. Jones. 2017. "Artificial Intelligence and Economic Growth." Working Paper No. 23928, National Bureau of Economic Research, October.

Agrawal, A., J.S. Gans, and A. Goldfarb. 2019. "Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction." *Journal of Economic Perspectives* 33(2):31–50.

Alekseev, A. 2020. "Give Me a Challenge or Give Me a Raise." ESI Working Paper No. 19-21, Chapman University, January.

—. 2019. "Using Response Times to Measure Ability on a Cognitive Task." *Journal of the Economic Science Association* 5:65–75.

Arrow, K.J. 1971. "The Economic Implications of Learning by Doing." *The Review of Economic Studies* 29:131–149.

Autor, D.H. 2015. "Why Are There Still So Many Jobs? The History and Future of Workplace Automation." *Journal of Economic Perspectives* 29(3):3–30.

Autor, D.H., F. Levy, and R.J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *The Quarterly Journal of Economics* 118:1279–1333.

Azrieli, Y., C.P. Chambers, and P.J. Healy. 2018. "Incentives in Experiments: A Theoretical Analysis." *Journal of Political Economy* 126:1472–1503.

Benabou, R., and J. Tirole. 2003. "Intrinsic and Extrinsic Motivation." *The Review of Economic Studies* 70:489–520.

Benndorf, V., H.A. Rau, and C. Sölch. 2018. "Minimizing Learning Behavior in Repeated Real-Effort Tasks." Working Paper No. 343, Center for European, Governance and Economic Development Research, Georg-August-Universität Göttingen.

Borghans, L., A.L. Duckworth, J.J. Heckman, and B. Ter Weel. 2008. "The Economics and Psychology of Personality Traits." *Journal of Human Resources* 43:972–1059.

Bradley, A.P. 1997. "The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms." *Pattern Recognition* 30:1145 – 1159.

Brynjolfsson, E., D. Rock, and C. Syverson. 2017. "Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics." Working Paper No. 24001, National Bureau of Economic Research.

Capra, C.M. 2019. "Understanding Decision Processes in Guessing Games: A Protocol Analysis Approach." *Journal of the Economic Science Association* 5:123–135.

Cason, T.N., L. Gangadharan, and N. Nikiforakis. 2011. "Can Real-Effort Investments Inhibit the Convergence of Experimental Markets?" *International Journal of Industrial Organization* 29:97–103.

Charness, G., D. Masclet, and M.C. Villeval. 2014. "The Dark Side of Competition for Status." *Management Science* 60:38–55.

Chung, S.H., Y.F. Su, and S.W. Su. 2012. "The Impact of Cognitive Flexibility on Resistance to Organizational Change." *Social Behavior and Personality* 40:735–745.

Cooper, D.J., I. Krajbich, and C.N. Noussair. 2019. "Choice-Process Data in Experimental Economics." *Journal of the Economic Science Association* 5:1–13.

Corgnet, B., and R. Hernán-González. 2019. "Revisiting the Trade-off Between Risk and Incentives: The Shocking Effect of Random Shocks?" *Management Science* 65:1096–1114.

Corgnet, B., R. Hernán-González, and R. Mateo. 2018. "Rac(g)e Against the Machine? Social Incentives When Humans Meet Robots." Working paper, University of Lyon.

Costa, P.T., and R.R. McCrae. 1992. *Revised NEO Personality Inventory (Neo-PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional manual*. Odessa, FL: Psychological Assessment Resources.

Coviello, D., A. Ichino, and N. Persico. 2014. "Time Allocation and Task Juggling." *American Economic Review* 104:609–23.

DeLong, E.R., D.M. DeLong, and D.L. Clarke-Pearson. 1988. "Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach." *Biometrics* 44:837–845.

Dewatripont, M., I. Jewitt, and J. Tirole. 1999. "The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government Agencies." *The Review of Economic Studies* 66:199–217.

Erkal, N., L. Gangadharan, and B.H. Koh. 2018. "Monetary and Non-Monetary Incentives in Real-Effort Tournaments." *European Economic Review* 101:528 – 545.

Ewers, M., and F. Zimmermann. 2015. "Image and Misreporting." *Journal of the European Economic Association* 13:363–380.

Falk, A., A. Becker, T. Dohmen, B. Enke, D. Huffman, and U. Sunde. 2018. "Global Evidence on Economic Preferences." *The Quarterly Journal of Economics* 133:1645–1692.

Falk, A., A. Becker, T.J. Dohmen, D. Huffman, and U. Sunde. 2016. "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences." Discussion Paper No. 9674, Institute for the Study of Labor (IZA).

Filiz-Ozbay, E., J.C. Ham, J.H. Kagel, and E.Y. Ozbay. 2018. "The Role of Cognitive Ability and Personality Traits for Men and Women in Gift Exchange Outcomes." *Experimental Economics* 21:650–672.

Fischbacher, U. 2007. "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics* 10:171–178.

Frederick, S. 2005. "Cognitive Reflection and Decision Making." *The Journal of Economic Perspectives* 19:25–42.

Frey, C.B., and M.A. Osborne. 2017. "The Future of Employment: How Susceptible are Jobs to Computerisation?" *Technological Forecasting and Social Change* 114:254–280.

Gill, D., and V. Prowse. 2012. "A Structural Analysis of Disappointment Aversion in a Real Effort Competition." *American Economic Review* 102:469–503.

Goos, M., A. Manning, and A. Salomons. 2014. "Explaining Job Polarization: Routine-Biased Technological Change and Offshoring." *American Economic Review* 104:2509–26.

Granulo, A., C. Fuchs, and S. Puntoni. 2019. "Psychological Reactions to Human versus Robotic Job Replacement." *Nature Human Behaviour* 3:1062–1069.

Hertwig, R., and I. Erev. 2009. "The Description–Experience Gap in Risky Choice." *Trends in Cognitive Sciences* 13:517–523.

Holmstrom, B., and P. Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics & Organization* 7:24–52.

Joy, S., E. Kaplan, and D. Fein. 2004. "Speed and Memory in the WAIS-III Digit Symbol-Coding Subtest Across the Adult Lifespan." *Archives of Clinical Neuropsychology* 19:759–767.

Kiesel, A., M. Steinhauser, M. Wendt, M. Falkenstein, K. Jost, A.M. Philipp, and I. Koch. 2010. "Control and Interference in Task Switching–A review." *Psychological Bulletin* 136:849–874.

Kohavi, R. 1995. "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection." In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*. vol. 2, pp. 1137–1143.

Krstajic, D., L.J. Buturovic, D.E. Leahy, and S. Thomas. 2014. "Cross-Validation Pitfalls when Selecting and Assessing Regression and Classification Models." *Journal of Cheminformatics* 6.

Lordan, G., and D. Neumark. 2018. "People Versus Machines: The Impact of Minimum Wages on Automatable Jobs." *Labour Economics* 52:40–53.

Loukidou, L., J. Loan-Clarke, and K. Daniels. 2009. "Boredom in the Workplace: More than Monotonous Tasks." *International Journal of Management Reviews* 11:381–405.

Martin, M.M., and R.B. Rubin. 1995. "A New Measure of Cognitive Flexibility." *Psychological Reports* 76:623–626.

Mokyr, J., C. Vickers, and N.L. Ziebarth. 2015. "The History of Technological Anxiety and the Future of Economic Growth: Is This Time Different?" *Journal of Economic Perspectives* 29(3):31–50.

Mullainathan, S., and J. Spiess. 2017. "Machine Learning: An Applied Econometric Approach." *Journal of Economic Perspectives* 31(2):87–106.

Muralidharan, K., M. Romero, and K. Wüthrich. 2019. "Factorial Designs, Model Selection, and (Incorrect) Inference in Randomized Experiments." Working Paper No. 26562, National Bureau of Economic Research, December.

Niederle, M., and L. Vesterlund. 2007. "Do Women Shy Away From Competition? Do Men Compete Too Much?" *The Quarterly Journal of Economics* 122:1067–1101.

Selten, R., D. Friedman, and A. Cassar. 2004. *Economics Lab: An Intensive Course in Experimental Economics*. Psychology Press.

Spector, A., and I. Biederman. 1976. "Mental Set and Mental Shift Revisited." *The American Journal of Psychology* 89:669–679.

Spiliopoulos, L., and A. Ortmann. 2018. "The BCD of Response Time Analysis in Experimental Economics." *Experimental Economics* 21:383–433.

Vandierendonck, A., B. Liefooghe, and F. Verbruggen. 2010. "Task Switching: Interplay of Reconfiguration and Interference Control." *Psychological Bulletin* 136:601–626.

Varian, H.R. 2014. "Big Data: New Tricks for Econometrics." *Journal of Economic Perspectives* 28(2):3–28.

Varma, S., and R. Simon. 2006. "Bias in Error Estimation when Using Cross-Validation for Model Selection." *BMC Bioinformatics* 7.

Wright, T.P. 1936. "Factors Affecting the Cost of Airplanes." *Journal of the Aeronautical Sciences* 3:122–128.

Zaniboni, S., D.M. Truxillo, and F. Fraccaroli. 2013. "Differential Effects of Task Variety and Skill Variety on Burnout and Turnover Intentions for Older and Younger Workers." *European Journal of Work and Organizational Psychology* 22:306–317.

Zeira, J. 1998. "Workers, Machines, and Economic Growth." *The Quarterly Journal of Economics* 113:1091–1117.

Appendices

A Proofs and Derivations

Proposition 1

Proof. The agent's utility in period 1 conditional on history up to that point is

$$U(e \mid x_1, h_1, \mu) = \mathbb{1}(e \geqslant \bar{e}_1(x_1))\pi(x_1 \mid \mu) - c_1(e).$$

Suppose the agent exerts effort $\bar{e}_1(x_1)$ and considers a positive deviation, Δe, from this level:

$$\begin{aligned} U(\bar{e}_1(x_1) + \Delta e \mid x_1, h_1, \mu) &= \pi(x_1 \mid \mu) - c_1(e + \Delta e) \\ &< \pi(x_1 \mid \mu) - c_1(e) \\ &= U(\bar{e}_1(x_1) \mid x_1, h_1, \mu). \end{aligned}$$

where the inequality follows since c is an increasing function. Hence, increasing effort beyond $\bar{e}_1(x_1)$ reduces the agent's utility. Next consider a negative deviation, $-\Delta e$, from $\bar{e}_1(x_1)$:

$$\begin{aligned} U(\bar{e}_1(x_1) - \Delta e \mid x_1, h_1, \mu) &= -c_1(e - \Delta e) \\ &< \pi(x_1 \mid \mu) - c_1(e) \\ &= U(\bar{e}_1(x_1) \mid x_1, h_1, \mu). \end{aligned}$$

where the inequality follows because the participation constraint holds. Hence, decreasing effort below $\bar{e}_1(x_1)$ reduces the agent's utility. These two inequalities imply that $e_1^* = \bar{e}_1(x_1)$ is the level of effort that maximizes the agent's utility in period 1. The agent's value in period 1 is then

$$U^*(x_1, h_1, \mu) = \pi(x_1 \mid \mu) - c_1(\bar{e}_1(x_1)).$$

Next consider the agent's total expected utility in period 0, given optimal behavior in period 1:

$$\begin{aligned} &U(e \mid x_0, h_0, \mu) + \mathbb{E}\, U^*(X_1, h_1, \mu) = \\ &= \mathbb{1}(e \geqslant \bar{e}_0(x_0))\pi(x_0 \mid \mu) - c_0(e) + \mathbb{E}\left[\pi(X_1 \mid \mu) - c_1(\bar{e}_1(X_1))\right], \end{aligned}$$

where X_1 denotes a random (from the perspective of period 0) task assigned in period 1. The distribution of X_1 conditional on a mode is given by equations (2) and (3). At $e = \bar{e}_0(x_0)$, the total utility is

$$\pi(x_0 \mid \mu) - c_0(\bar{e}_0(x_0)) + \mathbb{E}\,\pi(X_1 \mid \mu) - \mathbb{E}\left[c_1\left((\bar{e}_0(X_1) - \lambda \bar{e}_0(x_0)\,\mathbb{1}(H_1 = \eta_0))^+\right)\right],$$

where H_1 denotes a random type of a task in period 1 that has the same distribution as X_1. A positive deviation, Δe, from $\bar{e}_0(x_0)$ yields a total utility of

$$\pi(x_0 \mid \mu) - c_0(\bar{e}_0(x_0) + \Delta e) + \mathbb{E}\,\pi(X_1 \mid \mu) - \mathbb{E}\left[c_1\left((\bar{e}_0(X_1) - \lambda \bar{e}_0(x_0)\,\mathbb{1}(H_1 = \eta_0))^+\right)\right],$$

which is strictly less than the total utility at $\bar{e}_0(x_0)$ since the cost function is strictly increasing. Note that the cost of effort in period 1 does not change since the $\min\{\bar{e}_0(x_0) + \Delta e, \bar{e}_0(x_0)\}$ in the learning equation returns $\bar{e}_0(x_0)$. Hence, it is not optimal for the agent to exert more effort than required in period 0. Now consider a negative deviation, $-\Delta e$, from $\bar{e}_0(x_0)$. It yields a total utility

of

$$-c_0(\bar{e}_0(x_0) - \Delta e) + \mathbb{E}\,\pi(X_1 \mid \mu) - \mathbb{E}\left[c_1\left((\bar{e}_0(X_1) - \lambda \bar{e}_0(x_0)\,\mathbb{1}(H_1 = \eta_0) + \lambda \Delta e\,\mathbb{1}(H_1 = \eta_0)\right)^+\right]. \tag{A.1}$$

This total utility is strictly less than the total utility at $\bar{e}_0(x_0)$ since the participation constrain implies that

$$\pi(x_0 \mid \mu) - c_0(\bar{e}_0(x_0)) > -c_0(\bar{e}_0(x_0).$$

Additionally, the expected cost of effort in period 1 in (A.1) is less than the expected cost of effort in period 1 at $e_0 = \bar{e}_0(x_0)$. When there is no learning, the two terms coincide. When there is learning, however, a strict inequality holds. Hence, it is not optimal for the agent to exert less effort than required in period 0. Together these results imply that $e_0^* = \bar{e}_0(x_0)$ is the level of effort that maximizes the agent's total expected utility in period 0, conditional on optimal behavior in period 1. The agent's value in period 0 is then

$$U^*(x_0, h_0, \mu) = \pi(x_0 \mid \mu) - c_0(\bar{e}_0(x_0)).$$

□

Value Difference

Here I derive some of the expressions and results for the value difference between the Automatic and Manual modes, ΔV, for an arbitrary value of λ. Using the schemes of the Manual and Automatic modes, it is easy to derive the following expression for ΔV:

$$\begin{aligned}
\Delta V = {} & p\left[\pi^B - c_0(b) + c_0(a^l)\right] \\
& + p(1-p)\left[\pi^B - c_0(b) + c_0((a^l - \lambda a^h)^+) - s\right] \\
& + p(1-p)\left[c_0(a^h - \lambda a^l) - c_0(a^h) - s\right] \\
& + p^2\left[\pi^B - c_0(b - \lambda b) + c_0(a^l - \lambda a^l)\right].
\end{aligned}$$

Collecting the terms with π^B and $p(1-p)$, one obtains

$$\begin{aligned}
\Delta V = {} & 2p\pi^B + p\left[c_0(a^l) - c_0(b)\right] \\
& + p(1-p)\left[c_0((a^l - \lambda a^h)^+) + c_0(a^h - \lambda a^l) - c_0(b) - c_0(a^h) - 2s\right] \\
& + p^2\left[c_0(a^l - \lambda a^l) - c_0(b - \lambda b)\right] \\
= {} & p\{2\pi^B + c_0(a^l) - c_0(b) \\
& + (1-p)\left[c_0((a^l - \lambda a^h)^+) + c_0(a^h - \lambda a^l) - c_0(b) - c_0(a^h) - 2s\right] \\
& + p\left[c_0(a^l - \lambda a^l) - c_0(b - \lambda b)\right]\}
\end{aligned}$$

Denote

$$
\begin{aligned}
A &\equiv 2\pi^B + c_0(a^l) - c_0(b), \\
B &\equiv c_0(b) + c_0(a^h) - c_0((a^l - \lambda a^h)^+) - c_0(a^h - \lambda a^l) + 2s, \\
C &\equiv c_0(a^l - \lambda a^l) - c_0(b - \lambda b).
\end{aligned}
$$

Then the value difference can be written compactly as

$$\Delta V = (B + C)p\,(p - p^*)\,,$$

where

$$p^* = 1 - \frac{A + C}{B + C}. \tag{A.2}$$

It is easy to show that both $A + C$ and $B + C$ are strictly positive, hence $p^* < 1$. For p^* to be positive, however, one must have $A \leqslant B$, which after a little algebra becomes

$$\pi^B - c_0(b) \leqslant s + \frac{c_0(a^h) - c_0(a^l) - c_0((a^l - \lambda a^h)^+) - c_0(a^h - \lambda a^l)}{2}. \tag{A.3}$$

Then if condition (A.3) holds, the value difference will be negative for $p \leqslant p^*$ and positive for $p \geqslant p^*$. The value p^* serves as a unique threshold that separates the agent's choice of technology. This result makes sense intuitively: if automatability of the production process is high enough, the agent prefers the Automatic mode, and *vice versa*.

Consider condition (A.3). The LHS of it is the utility of task B net of the effort cost. The RHS of it is the sum of the switching cost and half of the term that I refer to as the *learning cost*, $l(\lambda)$. Condition (A.3) has an intuitive interpretation: the net utility of the new task has to be lower than the sum of the switching and learning costs. If the condition does not hold, i.e., the net utility of the new task is high enough, p^* will be negative and the agent will always prefer the Automatic mode.

It is worth noting that the learning cost is not guaranteed to be positive. For example, $l(0) = -2c_0(a^l) < 0$. On the other hand, $l(1) = c_0(a^h) - c_0(a^l) - c_0(\Delta a)$, which is positive because of the superadditivity of the cost function. In general, since l is a monotonically increasing function of λ, there will be a unique value of learning λ^*, such that $l(\lambda) < 0$ for $\lambda < \lambda^*$ and $l(\lambda) \geqslant 0$ for $\lambda \geqslant \lambda^*$. In other words, for the learning cost to be positive, the learning effect has to be strong enough. The threshold λ^* will depend on the effort requirements for tasks Al and Ah, as well as on the shape of the cost function.

Using these results, it is now easy to prove Proposition 2.

Proposition 2

Proof. Consider formula (A.2). Putting $\lambda = 1$, one obtains

$$
\begin{aligned}
B &= c_0(b) + c_0(a^h) - c_0(\Delta a) + 2s, \\
C &= 0,
\end{aligned}
$$

which implies that

$$p^* = 1 - \frac{2\pi^B + c_0(a^l) - c_0(b)}{c_0(b) + c_0(a^h) - c_0(\Delta a) + 2s}.$$

Putting $\lambda = 1$ in condition (A.3) yields

$$\pi^B - c_0(b) \leqslant s + \frac{c_0(a^h) - c_0(a^l) - c_0(\Delta a)}{2}.$$

□

Proposition 3

Proof. Assume that $p \in (0, 1)$. Denote

$$B' \equiv c_0(b) + c_0(a^h) - c_0((a^l - \lambda a^h)^+) - c_0(a^h - \lambda a^l)$$

The probability that agent i chooses the Automatic mode is

$$\begin{aligned}
\mathbb{P}(y_i = 1) &= \mathbb{P}(\Delta V_i \geqslant 0) \\
&= \mathbb{P}\left(p\left[A - (1-p)(B' + 2s_i) + pC\right] \geqslant 0\right) \\
&= \mathbb{P}\left(A - (1-p)B' - 2(1-p)s_i + pC \geqslant 0\right) \\
&= \mathbb{P}\left(2(1-p)s_i \leqslant A - (1-p)B' + pC\right) \\
&= \mathbb{P}\left(2s_0 + 2\sigma\epsilon_i \leqslant \frac{A}{1-p} - B' + \frac{p}{1-p}C\right) \\
&= \mathbb{P}\left(\epsilon_i \leqslant \frac{1}{2\sigma}\left[\frac{A}{1-p} - B' + \frac{p}{1-p}C - 2s_0\right]\right) \\
&= \Phi\left(\frac{1}{2\sigma}\left[\frac{A}{1-p} - B' + \frac{p}{1-p}C - 2s_0\right]\right).
\end{aligned}$$

Putting $\lambda = 1$ yields

$$\begin{aligned}
B' &= c_0(b) + c_0(a^h) - c_0(\Delta a) \\
C &= 0,
\end{aligned}$$

which implies that

$$\mathbb{P}(y_i = 1) = \Phi\left(\frac{1}{2\sigma}\left[\frac{2\pi^B + c_0(a^l) - c_0(b)}{1-p} + c_0(\Delta a) - c_0(b) - c_0(a^h) - 2s_0\right]\right).$$

If $p = 0$, we have

$$\mathbb{P}(y_i = 1) = \mathbb{P}(\Delta V_i \geqslant 0) = \mathbb{P}(0 \geqslant 0) = 1.$$

If $p = 1$, we have

$$\mathbb{P}(y_i = 1) = \mathbb{P}(\Delta V_i \geqslant 0) = \mathbb{P}(A + C \geqslant 0) = 1.$$

An alternative expression for the probability of choosing the Automatic mode can be obtained as follows. Denote

$$B_0 \equiv c_0(b) + c_0(a^h) - c_0((a^l - \lambda a^h)^+) - c_0(a^h - \lambda a^l) + 2s_0.$$

Then

$$\begin{aligned}\mathbb{P}(y_i = 1) &= \Phi\left(\frac{1}{2\sigma}\left[\frac{A}{1-p} + \frac{p}{1-p}C - B_0\right]\right) \\ &= \Phi\left(\frac{1}{2\sigma(1-p)}\left[A + pC - (1-p)B_0\right]\right) \\ &= \Phi\left(\frac{1}{2\sigma(1-p)}\left[A - B_0 + p(B_0 + C)\right]\right) \\ &= \Phi\left(\frac{B_0 + C}{2\sigma(1-p)}(p - p_0^*)\right),\end{aligned}$$

where

$$p_0^* \equiv 1 - \frac{A + C}{B_0 + C}.$$

Putting $\lambda = 1$ yields

$$\Phi\left(\frac{(c_0(a^h) + c_0(b) - c_0(\Delta a) + 2s_0)(p - p_0^*)}{2\sigma(1-p)}\right),$$

and

$$p_0^* = 1 - \frac{2\pi^B - c_0(b) + c_0(a^l)}{c_0(a^h) + c_0(b) - c_0(\Delta a) + 2s_0}.$$

□

Proposition 4

Proof. Let $D \equiv \frac{1}{2\sigma}\left(\frac{A}{1-p} - B' + \frac{p}{1-p}C - 2s_0\right)$.

1. Consider the partial derivative of $d_{\mathcal{A}}$ w.r.t. p

$$\frac{\partial d_{\mathcal{A}}}{\partial p} = \phi(D)\frac{A + C}{2\sigma(1-p)^2} > 0.$$

2. Consider the partial derivative of $d_{\mathcal{A}}$ w.r.t. s_0

$$\frac{\partial d_{\mathcal{A}}}{\partial s_0} = -\frac{\phi(D)}{\sigma} < 0.$$

3. Consider the partial derivative of $d_{\mathcal{A}}$ w.r.t. σ

$$\frac{\partial d_{\mathcal{A}}}{\partial \sigma} = -\frac{\phi(D)}{\sigma^2}\frac{(B_0 + C)(p - p_0^*)}{2(1-p)} \begin{cases} \leqslant 0, & p \geqslant p_0^*, \\ > 0, & p < p_0^*. \end{cases}$$

4. Consider the partial derivative of $d_{\mathcal{A}}$ w.r.t. λ

$$\begin{aligned}\frac{\partial d_{\mathcal{A}}}{\partial \lambda} = \frac{\phi(D)}{2\sigma}[&-a^h c_0'(a^l - \lambda a^h)\,\mathbb{1}(a^l \geqslant \lambda a^h) - a^l c_0'(a^h - \lambda a^l) \\ &- \frac{p}{1-p}a^l c_0'(a^l - \lambda a^l) + \frac{p}{1-p}b c_0'(b - \lambda b)].\end{aligned}$$

If condition

$$\frac{p}{1-p}bc_0'(b-\lambda b) \geqslant a^h c_0'(a^l-\lambda a^h)\,\mathbb{1}(a^l \geqslant \lambda a^h) + a^l c_0'(a^h-\lambda a^l) + \frac{p}{1-p}a^l c_0'(a^l-\lambda a^l)$$

is satisfied, then

$$\frac{\partial d_{\mathcal{A}}}{\partial \lambda} \geqslant 0.$$

5. Consider the partial derivative of $d_{\mathcal{A}}$ w.r.t. π^B

$$\frac{\partial d_{\mathcal{A}}}{\partial \pi^B} = \frac{\phi(D)}{\sigma(1-p)} > 0.$$

6. Consider the partial derivative of $d_{\mathcal{A}}$ w.r.t. b

$$\frac{\partial d_{\mathcal{A}}}{\partial b} = -\frac{\phi(D)}{2\sigma}\left(\frac{c_0'(b)}{1-p} + c_0'(b) + \frac{p}{1-p}(1-\lambda)c_0'(b-\lambda b)\right) < 0.$$

7. Consider the partial derivative of $d_{\mathcal{A}}$ w.r.t. a^h

$$\frac{\partial d_{\mathcal{A}}}{\partial a^h} = -\frac{\phi(D)}{2\sigma}\left(\lambda c_0'(a^l-\lambda a^h)\,\mathbb{1}(a^l \geqslant \lambda a^h) + c_0'(a^h) - c_0'(a^h-\lambda a^l)\right) < 0,$$

since c_0' is an increasing function.

8. Consider the partial derivative of $d_{\mathcal{A}}$ w.r.t. a^l

$$\frac{\partial d_{\mathcal{A}}}{\partial a^l} = \frac{\phi(D)}{2\sigma}\left(\frac{c_0'(a^l)}{1-p} + c_0'(a^l-\lambda a^h)\,\mathbb{1}(a^l \geqslant \lambda a^h) + \frac{p}{1-p}(1-\lambda)c_0'(a^l-\lambda a^l) - \lambda c_0'(a^h-\lambda a^l)\right).$$

If condition

$$\frac{c_0'(a^l)}{1-p} + c_0'(a^l-\lambda a^h)\,\mathbb{1}(a^l \geqslant \lambda a^h) + \frac{p}{1-p}(1-\lambda)c_0'(a^l-\lambda a^l) \geqslant \lambda c_0'(a^h-\lambda a^l),$$

is satisfied, then

$$\frac{\partial d_{\mathcal{A}}}{\partial a^l} \geqslant 0.$$

□

B Subject Instructions

Note that subjects instruction were delivered on screen and [...] elements were replaced by actual values generated by the software.

Introduction

Welcome and thank you for participating! This is an experiment in individual economic decision-making. Please, mute/turn off all of your electronic devices for the duration of the experiment.

Session overview

Today's session will consist of two parts, Part A and Part B. The session will take up no more than 2 hours. You will proceed with the experiment at your own pace.

Payment

Your total payment will consist of a participation payment of $[...], a payment for the experimental tasks, and a payment for the questionnaire. You will be paid privately in cash at the end of the experiment.

Privacy

You will not interact with other participants. Please, do not reveal your identity to anyone. You must not talk to other participants during the experiment.

Final Notes

Please read the following instructions carefully. You are welcome to ask questions at any point. Just raise your hand and we will answer your questions in private.

Structure of experiment

The experiment consists of Part A and Part B. Part A of the experiment is split into three Phases. In the first Phase, you will encounter the so-called [...] mode. In the second Phase, you will encounter the so-called [...] mode. In the third Phase, you will first be given an option to make a choice between the two modes. This choice will determine under which mode you will work in the third Phase. The meaning of the two modes will be explained shortly.

Manual mode

In the Manual mode, you will work on the Encryption task. In the Encryption task, you will need to find correct matches between letters and numbers. In each round, you will see 18 pairs of letters and numbers arranged in a table (the Key) at the upper part of the screen. Below the Key, there will be a smaller table with letters and empty input boxes (the Lock). Your goal is to fill in the boxes in the Lock with numbers according to the Key. You will fill in the boxes by typing numbers on your keyboard.

The Lock will contain either 3 (3-item Lock) or 6 (6-item Lock) empty boxes. The computer will decide before the start of each round whether you will be given a 3-item or a 6-item Lock.

The chance that you will be given a 6-item Lock in any given round is [...]%. This means that, on average, every [...] out of five rounds will have a 6-item Lock.

After filling in the boxes as you see fit, you will need to click Submit to proceed to the next round. You will complete the task at your own pace. There will be [...] rounds of the task. A round will be considered as correctly solved if all the boxes in the Lock are filled in according to the Key. You will learn your score at the end of the experiment. You will receive $[...] for each correctly solved round.

Automatic mode

In the Automatic mode, the computer will help you with the Encryption task. The computer will be able to help you with the 3-item Lock, but not with the 6-item Lock. Here is how it works. As in the Manual mode, the computer will decide before the start of each round whether the round will have a 3-item or a 6-item Lock. If the round turns out to have a 6-item Lock (recall that there is a [...]% chance of that happening), you will work on the Encryption task with a 6-item Lock, as before. However, if the round turns out to have a 3-item Lock, the computer will correctly solve this round for you. It will take [...] seconds for the computer to correctly solve the round. You, in the meantime, will have a chance to work on the Counting task.

In the Counting task, you will need to count the number of happy faces in a table of [...] happy and frowny faces. You will enter your answer in the empty input box by typing the number using your keyboard. After filling in the box as you see fit, you will need to click Submit to proceed to the next round. If you click Submit before the computer ends working on a 3-item Lock, you will proceed to a new round of the Counting task. If the computer finishes working on a 3-item Lock before you click Submit in the Counting task, one of the two things will happen.

First of all, the computer will decide whether the next round of the Encryption task will have a 3-item or a 6-item Lock. If the round turns out to have a 3-item Lock, the computer will proceed with that round, and you will continue working on your current Counting task. If, however, the round turns out to have a 6-item Lock, the computer will have to interrupt your work on the Counting task and ask you to solve the 6-item Lock.

There will be [...] rounds of the Encryption task, during which you can work. You will be able to complete as many rounds of the Counting task as you can, while the computer works on 3-item Locks. The Automatic mode will finish after the [...]th round of the Encryption task is finished either by you or the computer.

You will learn your score at the end of the experiment. You will receive $[...] for each correctly solved round of the Counting task, $[...] for each correctly solved round of 6-item Encryption task, and $[...] for each round of a 3-item Encryption task (the computer always solves these rounds correctly for you).

Practice

You will have [...] practice rounds of each task before Phase 1 begins. This will give you a chance to familiarize yourself with the interfaces of the tasks. During the practice, you will receive feedback on your performance. You will receive no feedback on your performance in the further rounds.

Choice Phase

At the beginning of the third Phase, you will be given an option to choose under which mode you would like to work for the last Phase of Part A. If you choose the Manual mode, you will work

according to the rules of the Manual mode for [...] rounds. If you choose the Automatic mode, you will work according to the rules of the Automatic mode for [...] rounds.

Part B

After finishing the third Phase, you will proceed to Part B of the experiment. In Part B, you will be given a questionnaire that consists of several parts. You will be asked to answer the questions in the questionnaire to proceed.

End of experiment and payment

After you complete the questionnaire, you will see your results from all the Phases: the Phase with the Manual mode, the Phase with the Automatic mode, and the Phase with the mode of your choice. The computer will pick at random one Phase out of those three for the final payment. Each Phase will be equally likely to be picked by the computer. You will not know which Phase determines your earnings for the experiment when you go through the three Phases. It makes sense, therefore, to perceive each Phase as if it were the one that determines your earnings. The last screen will show you which Phase was chosen for your payment by the computer, and your earnings in the experiment.

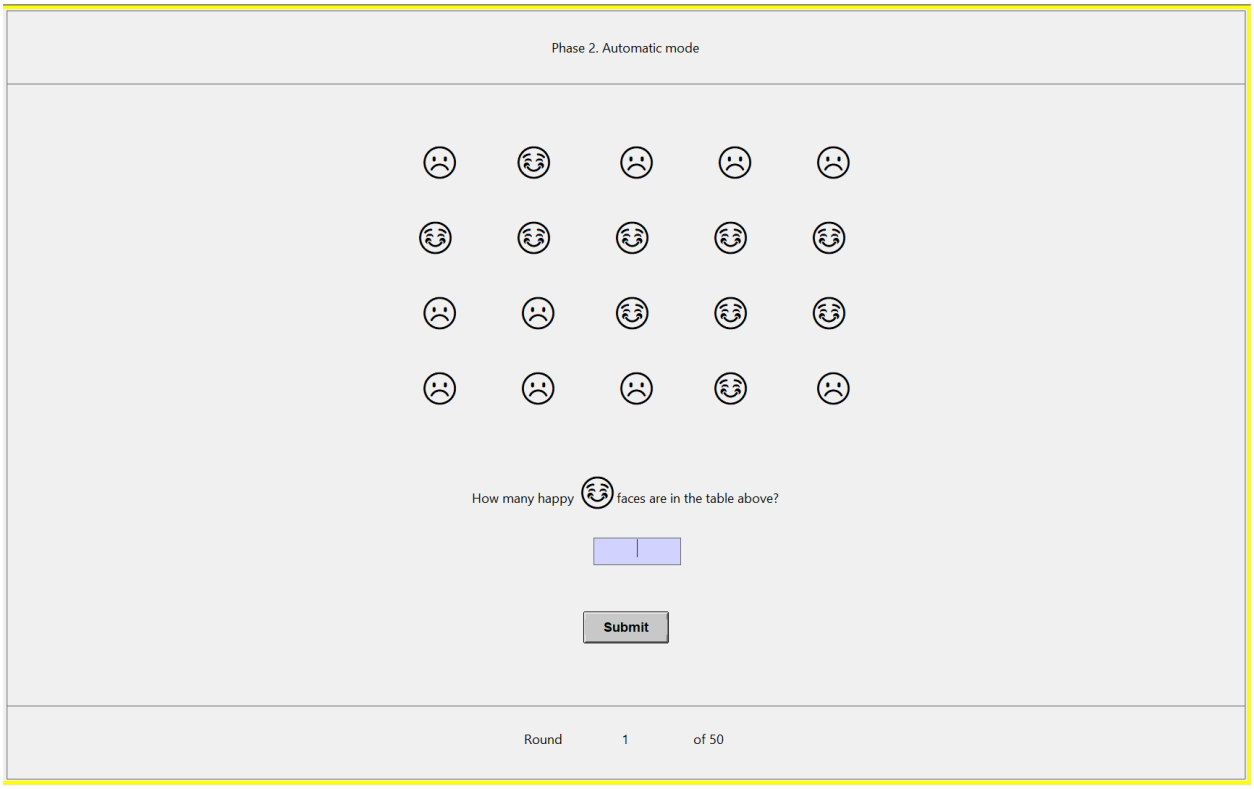
If you have any questions about the experiment, please raise your hand, and we will help you in private. If you are ready to proceed to the task, click Proceed to begin the practice rounds.

C Screenshots

Figure C.1: Encryption Task in the Manual Mode

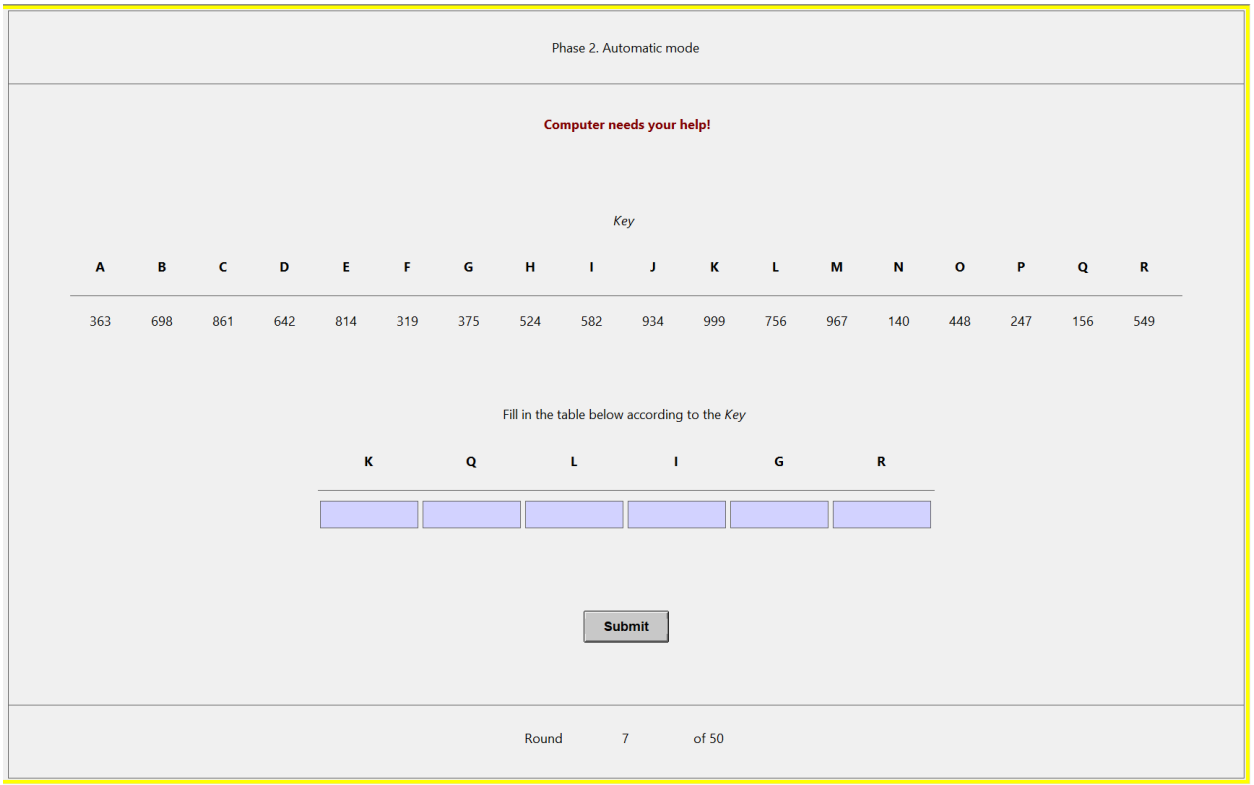
Note: The figure shows an example of an Encryption task with 3 items in the Manual mode.

Figure C.2: Counting Task



Note: The figure shows an example of a Counting task in the Automatic mode.

Figure C.3: Encryption Task in the Automatic Mode



Note: The figure shows an example of an Encryption task in the Automatic mode. The Encryption task in the Automatic mode always had 6 items. The message at the top of the screen indicates that subjects' previous work on the Counting task was interrupted.

D Additional Tables

Table D.1: Demographic Characteristics of the Sample

Characteristic	Mean
Gender and Age	
Female	0.58
Male	0.42
Age	21.34
Race	
Asian	0.21
Black	0.01
Hispanic	0.15
Other	0.08
White	0.55
Year in School	
Freshman	0.03
Sophomore	0.10
Junior	0.26
Senior	0.60
Graduate	0.01
Major	
Arts	0.36
Business	0.21
Economics	0.04
Other	0.21
STEM	0.19

E Textual Analysis

The notes below explain the classification criteria and sample answers for each reason category.

Manual Mode

- dislike smiley face: Subjects do not like smiley face. Subject 81: "Looking at the smiley/frowny faces felt like it was tricking my mind."
- enough time: Subjects think they have enough time to finish the task. Subject 90: "Because I felt very rushed during the Counting part for the Automatic mode. I felt like it was better for me to feel calm and composed while doing the Manual mode and take my time on the coding part."
- easy focus: Subjects think it is not easy to be distracted in this mode. Subject 119: "I found it easier to focus on one task as opposed to switching between two different ones."
- more confident: Subjects think that they can do it more correctly. Subject 127: "Know I can do it right."
- more efficient: Subjects think that it is more efficient under this mode. Subject 86: "I felt I was more efficient in just looking for the letter and copying the number."
- more control: subjects think that they have more control in this mode. Subject 83: "I knew I missed a round or two in automatic, whereas I had more control over the rounds in manual. I knew I could get all the rounds correct in manual."
- thinking consistency: Subjects do not want to change their mind between different modes. Subject 89: "I didn't want to switch between thinking modes."

Automatic Mode

- challenge themselves: Even realizing that it is hard, subjects want to challenge themselves. Subject 155: "I chose the Automatic mode because I thought it was more challenging to switch between encryption task and counting task. I wanted to challenge myself and see how many I could still get right, even if it was more of a headache"
- earn more: Subjects think they can earn more money. Subject 112: "Although it was harder, you are guaranteed more money".
- like smiley face: Subjects like to count smiley face. Subject 118 "I preferred counting smiley faces over inputting encryptions."
- much faster: Subjects think it's much faster in this mode. Subject 140: "It's faster to count the faces than inputting numbers"
- more correct: Subjects think some tasks in this mode will be more correct if it is done by computers. Subject 138 "I trusted the computer more than myself."
- not monotonous: Subjects think automatic mode is not as tedious as manual mode. Subject 135: "Typing in everything manually got very tedious after a while. When the smiley faces came up, it gave me a brain break and it felt like a game."